



PUBLIC

SAP Data Services

Document Version: 4.2 Support Package 12 (14.2.12.0) – 2020-02-06

Text Data Processing Language Reference Guide

Content

- 1 Text Data Processing Language Reference Guide. 4**
- 2 Overview of the Language Reference Guide. 5**
 - 2.1 Who Should Read This Guide. 5
- 3 Overview of Text Data Processing Functionality. 6**
 - 3.1 Character Encodings. 9
 - 3.2 About Linguistic Analysis. 9
 - 3.3 About Entity Extraction. 9
 - About Customizing Extraction. 10
 - 3.4 About Fact Extraction. 10
 - About Sentiment Analysis 11
 - About Public Sector Fact Extraction. 11
 - About Enterprise Fact Extraction. 11
- 4 Linguistic Analysis 13**
 - 4.1 Word Segmentation. 14
 - Whitespace Languages. 14
 - Nonwhitespace Languages. 30
 - 4.2 Case Normalization. 39
 - 4.3 Stemming. 40
 - Language-Specific Stemming Examples. 41
 - Stemming Unknown Words. 81
 - 4.4 Compound Analysis. 81
 - Language-Specific Compound Analysis Examples. 81
 - 4.5 Part-of-Speech Tagging. 87
 - Part-of-Speech Naming Conventions. 88
 - Part-of-Speech and Stem Disambiguation 88
 - Guessing Part-of-Speech for Unknown Words. 89
 - Language-Specific Unknown Words Guessing Examples. 89
 - Language-Specific Part-of-Speech Tagging Examples. 93
- 5 Entity Extraction. 208**
 - 5.1 Predefined Entity Types by Language. 208
 - 5.2 Language-specific Entity Type Examples. 214
 - 5.3 Definitions of Entity Types. 295
 - 5.4 Entity Normalization. 308
- 6 Fact Extraction. 314**

6.1	Sentiment Analysis Fact Extraction.	314
	Sentiment Extraction.	315
	Emoticon Extraction.	329
	Request Extraction.	340
	Profanity Extraction.	346
6.2	Public Sector Fact Extraction.	346
	Common Entity Extraction.	348
	Action Event Extraction.	353
	Military Unit Extraction.	355
	Organizational Information Extraction.	356
	Person: Alias Extraction.	357
	Person: Appearance Extraction.	359
	Person: Attribute Extraction.	361
	Person: Relationship Extraction.	362
	Spatial Reference Extraction.	364
	Travel Event Extraction.	365
6.3	Enterprise Fact Extraction.	367
	Management Change Event Extraction.	368
	Membership Information Extraction.	371
	Merger Information Extraction.	373
	Organizational Information Extraction.	374
	Product Release Event Extraction.	376
6.4	Entity Type Names and Subentity Type Names For Fact Extraction.	378

1 Text Data Processing Language Reference Guide

Text Data Processing provides a suite of analyzers that enable developers to construct applications that analyze unstructured data, enabling your application to perform extraction and natural-language processing of text. This guide provides information about the linguistic extraction features, and describes the behavior of the supported language modules.

2 Overview of the Language Reference Guide

Welcome to the *Language Reference Guide*.

This guide addresses the variety of languages that can be processed by the text analytics engine: a suite of natural-language processing capabilities based on linguistic, statistical and machine-learning algorithms that model and structure the information content of textual sources in multiple languages.

This technology forms the foundation for advanced text processing for a range of applications including search, business intelligence or exploratory data analysis.

The major areas of text analytics functionality are:

Linguistic Analysis is the most fundamental form of text analytics - tokenization, stemming and part-of-speech tagging. These operations allow for optimized full-text index building because they guarantee both high precision and recall.

Entity Extraction is the identification of named entities (persons, organizations, and so on), which eliminates the 'noise' in textual data by highlighting salient information. This process transforms unstructured text into structured information.

Fact Extraction is a higher-level semantic processing that links entities as "facts" in domain-specific applications. For example, "Voice of the Customer" classifies sentiments with their corresponding topics.

2.1 Who Should Read This Guide

Users of this guide may need to enhance extraction in their text analytics application and should understand text data processing extraction concepts. However, users of this guide are not expected to understand or be familiar with the natural languages of the text being processed by the software. Similarly, users are not required to be familiar with linguistic principles. This document assumes the following:

- You are an application developer or consultant working on enhancing text data processing extraction.
- You understand your organization's text data processing extraction needs.

3 Overview of Text Data Processing Functionality

The software includes language modules for the languages supported. Each language module consists of a set of files that include system dictionaries containing words to support the language processing operations for the given natural language.

It is the language modules that enable linguistic analysis and extraction of unstructured text in a given language. Language modules use the following language processing technologies:

- Linguistic analysis to handle natural language processing
- Entity extraction to handle named entity extraction and normalization
- Fact extraction to handle sentiment analysis, public sector events, and enterprise facts

Linguistic analysis provides

- Word segmentation
- Stemming
- Part of speech tagging
- Tagged stemming (part of speech in context)

Extraction provides the following levels:

- Entity Extraction
 - Core Extraction – for named entities (PERSON, DATE, LOCALITY, and so on.)
- Fact Extraction
 - Sentiment Analysis – for positive and negative sentiment and topic extraction
 - Public Sector - for public-sector-specific information.
 - Enterprise - for business-specific information

There is an additional language module for the "neutral" language. The neutral language module offers basic tokenization, stemming and part-of-speech tagging in any language other than the 33 currently covered by SAP Text Analysis. The neutral language works for whitespace languages, such as Afrikaans or Bengali. Customization is also possible through dictionaries and rules. Refer to the *Extraction Customization Guide* for more details about customization.

The following table shows the languages which are explicitly supported for Text Data Processing.

Language	Linguistic Analysis	Entity Extraction			Fact Extraction		
		Predefined	Custom Dictionary	Sentiment Analysis	Public Sector	Enterprise	Custom Rules
Arabic (AR)	X	X	X	X			X
Catalan (CA)	X		X				X
Chinese, simplified (ZH)	X	X	X*	X			X

Language	Linguistic Analysis	Entity Extraction			Fact Extraction		
		Predefined	Custom Dictionary	Sentiment Analysis	Public Sector	Enterprise	Custom Rules
Chinese, traditional (ZF)	X	X	X*	X			X
Croatian (HR)	X		X				X
Czech (CS)	X		X				X
Danish (DA)	X		X				X
Dutch (NL)	X	X	X	X**			X
English (EN)	X	X	X	X	X	X	X
Farsi (FA)	X	X	X				X
French (FR)	X	X	X	X			X
German (DE)	X	X	X	X			X
Greek (EL)	X		X				X
Hebrew (HE)	X		X				X
Hungarian (HU)	X		X				X
Indonesian (ID)	X		X				X
Italian (IT)	X	X	X	X			X
Japanese (JA)	X	X	X*	X			X
Korean (KO)	X	X	X				X
"Neutral" language (UD)	X		X				X
Norwegian Bokmal (NB)	X		X				X
Norwegian Nynorsk (NN)	X		X				X
Polish (PL)	X		X				X

Language	Linguistic Analysis	Entity Extraction			Fact Extraction		
		Predefined	Custom Dictionary	Sentiment Analysis	Public Sector	Enterprise	Custom Rules
Portuguese (PT)	X	X	X	X			X
Romanian (RO)	X		X				X
Russian (RU)	X	X	X	X			X
Serbian Cyrillic (SR)	X		X				X
Serbian Latin (SH)	X		X				X
Slovak (SK)	X		X				X
Slovenian (SL)	X		X				X
Spanish (ES)	X	X	X	X			X
Swedish (SV)	X		X				X
Thai (TH)	X		X*				X
Turkish (TR)	X		X				X

* Dictionary matching in CCJT languages operates at the character level; i.e., each character is treated as a separate token. For more details about this, refer to the *Extraction Customization Guide*.

** Emoticon and profanity extraction only.

Related Information

[Linguistic Analysis \[page 13\]](#)

[Entity Extraction \[page 208\]](#)

[Fact Extraction \[page 314\]](#)

3.1 Character Encodings

Table of encodings accepted by the SAP text analytics engine.

Unicode	ISO IEC	Code Page	Other
utf_8	iso_8859_1	cp_1250	bog5
utf_16	iso_8859_2	cp_1251	euc_cn
ucs_4	iso_8859_5	cp_1252	euc_jp
	iso_8859_7	cp_1253	euc_kr
	iso_8859_9	cp_1256	gb_18030
			gb_2312_80
			koi8_r
			shift_jis
			tis_620

3.2 About Linguistic Analysis

The software provides and uses sophisticated natural language processing capabilities for linguistic analysis of unstructured data. Some of these capabilities include:

- Segmentation—the separation of input text into its elements
- Stemming—the identification of word stems, or dictionary forms
- Tagging—the labeling of words' parts of speech

Related Information

[Linguistic Analysis \[page 13\]](#)

[Word Segmentation \[page 14\]](#)

[Stemming \[page 40\]](#)

[Part-of-Speech Tagging \[page 87\]](#)

3.3 About Entity Extraction

Entity extraction is the process of finding and marking specific named entities that occur in unstructured text.

Entities denote the names of people, places, things, dates, values, and so forth, that can be extracted from text. An entity is defined as a pairing of a standard form and its type. For example, `Winston Churchill`/`PERSON` is an entity in which `Winston Churchill` is the standard form and `PERSON` is the type.

The language modules included with the software contain system dictionaries and provide an extensive set of predefined entity types. The extraction process can extract entities using these lists of specific entities. Extraction classifies each extracted entity by entity type and presents this metadata in a standardized format.

Related Information

[Entity Extraction \[page 208\]](#)

3.3.1 About Customizing Extraction

Introduces the concepts of creating custom dictionaries and extraction rules that can be created to extend the capabilities of the software.

You can enhance the extraction process by creating and using:

- Dictionaries that contain information about entities. You can customize information about the entities your application must find.
- Extraction rules.

For details about enhancing extraction, refer to the *SAP Data Services Text Data Processing Extraction Customization Guide*.

For certain language modules, you can also enhance extraction by using the specialized extraction content included in them.

3.4 About Fact Extraction

Fact extraction is the process of discovering and presenting facts about or relationships between entities that occur in unstructured text, for example the sentiment expressed about a particular product or a merger event between two companies.

Related Information

[Fact Extraction \[page 314\]](#)

3.4.1 About Sentiment Analysis

Sentiment Analysis (sometimes also referred to as "Voice of the Customer") is the process of using rules to retrieve specific information about customers' sentiments and requests when processing and analyzing text. These same rules also retrieve emoticons and profanities.

Positive and negative opinions, such as "great" or "awful" are extracted as sentiments.

For example, "I'm so excited about this car!" is extracted as a Strong Positive Sentiment.

Related Information

[Sentiment Analysis Fact Extraction \[page 314\]](#)

3.4.2 About Public Sector Fact Extraction

Public Sector facts are extracted using rules consisting of patterns that define the expressions used to represent events relevant for the public sector.

Events like travel and the people and destinations involved in them are extracted as public sector facts. For example, "John Doe was arrested on Christmas Day, when he arrived in San Francisco from England." is a public sector fact extracted as a Travel Event.

Related Information

[Public Sector Fact Extraction \[page 346\]](#)

3.4.3 About Enterprise Fact Extraction

Enterprise facts are extracted using rules consisting of patterns that define domain specific expressions used to represent relationships and events relevant for an enterprise.

Events like mergers and acquisitions and management changes are extracted as enterprise facts. For example, "New Oji Paper and Honshu Paper Co Ltd merged on Tuesday to become the third-largest paper and pulp producer in the world." is an enterprise fact extracted as a Merge Event.

Related Information

[Enterprise Fact Extraction \[page 367\]](#)

4 Linguistic Analysis

The software provides and uses many features for multilingual natural language processing (NLP) of unstructured data. Among these are language coding and identification, segment generation, word segmentation, stemming, and part-of-speech tagging.

Feature	Description
Language and encoding identification	The automatic recognition of the input language, for example, French or Japanese, and of various character encodings (such as Unicode UTF-8 and Code Page 1252).
Segment generation	The breaking of input text into segments of one or more complete paragraphs and sentences for more efficient processing.
Word segmentation	The separation of input text into its elements, such as words and punctuation.
Case normalization	The normalization of the initial letter of a word to upper or lower case. Used to counteract "accidental" case changes for words that appear in titles, headings, or at the beginning of sentences.
Stemming	The identification of word stems, or dictionary forms, for text or single words.
Compound analysis	The separation of compound words into their component stems.
Part-of-speech tagging	The labeling of words' parts of speech, for example, noun or verb. Where stemming can produce multiple results (ambiguity), part-of-speech tagging selects one result, based on the surrounding context.
Document analysis	The recognition of a document's major sections—paragraphs and sentences.
Tagged stemming	The identification of word stems for a word of a given part-of-speech.

i Note

Not all operations are supported for all languages.

Related Information

[Word Segmentation \[page 14\]](#)

[Stemming \[page 40\]](#)

[Compound Analysis \[page 81\]](#)

[Part-of-Speech Tagging \[page 87\]](#)

[Case Normalization \[page 39\]](#)

4.1 Word Segmentation

Breaking text into the smallest, meaningful syntactic units, such as words or punctuation.

The word segmentation operation performs basic word breaking. The word segmenter also identifies idiomatic phrases, such as **case in point** or **out-of-the-box**. These idiomatic phrases are processed as a single unit or word. Hyphenated words are not broken, since they are syntactic units. However, contractions (such as **don't**) and elisions (such as **l'abri**) are separated into their syntactic units.

Related Information

[Multiword Units \[page 14\]](#)

4.1.1 Whitespace Languages

Whitespace languages mark word boundaries with whitespace and punctuation marks.

This group includes European, Balkan, and Middle Eastern languages, as well as Korean and the Neutral language. Punctuation marks sometimes end a sentence, in which case they are used in sentence detection.

Alphanumeric expressions referring to a value and a measurement unit (55Kg, 143.5Km/h) are split in most whitespace languages (Bokmal, Catalan, Danish, Dutch, English, French, German, Italian, Nynorsk, Portuguese, Spanish, Swedish, Neutral). For example, the input text "55Kg" is segmented as "55 Kg", and "143.5Km/h" becomes "143.5 Km/h" after segmentation.

i Note

The "neutral" language module offers basic segmentation, stemming and part-of-speech tagging in any whitespace language not already fully covered by Text Data Processing, such as Chechen or Somali.

4.1.1.1 Multiword Units

By default, multiword units are segmented as a single unit, for example, "to and fro " and "Buenos Aires" are each segmented as one unit. However, you can turn this behavior off, and in such a case, multiword units will be broken down into their individual components. For example, "to and fro" is segmented into three units instead of one.

4.1.1.2 Punctuation

Word segmentors generally split off punctuation marks as separate units. This includes periods and commas, sentence-ending punctuation, and various quotation marks.

The following table summarizes punctuation-related segmentation conventions:

No Whitespace	If a punctuation mark is followed by a character and not by whitespace, it is not split off from its surrounding word. For example: "filename.filetype" is segmented as "filename.file-type".
Abbreviations	Abbreviations ending in a period are important exceptions to the general rule that splits punctuation from their terms; their periods remain with them.
Apostrophes	Contractions spelled with apostrophes (like can't, don't, and so on in English) are handled by language-specific rules.
Hyphens	Embedded and trailing hyphens are not split off from their words. Leading hyphens are not split off before a digit expression, for example, -1000 is segmented as one unit.

4.1.1.3 Language-Specific Whitespace Segmentation Examples

A collection of examples of word segmentation for the whitespace languages.

4.1.1.3.1 Arabic Word Segmentation

The Arabic segmenter follows all of the general segmentation rules in the whitespace languages. Specifically, words are broken at whitespace and punctuation. The only exceptions to this are for punctuation embedded inside a word (for example, **www.microsoft.command** and **1,500**), and space embedded in masculine, feminine, or isolated forms of cardinal and ordinal numbers, such as **المائة وأحد عشرة**.

4.1.1.3.2 Catalan Word Segmentation

The Catalan segmenter follows all of the general segmentation rules in the whitespace languages, but it has certain language-specific behavior.

Contracted forms, such as **pel** are not split. Elided pronouns preceding verbs are separated from the verb, as is the case with **m'** in **m'han vist**. Clitics appearing after the verb are split, both in their reduced (**'l** in **posa'l**) or full form (**-lo** in **posin-lo**).

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.1.3.3 Croatian Word Segmentation

The Croatian segmenter follows all of the general segmentation rules in the whitespace languages.

4.1.1.3.4 Czech Word Segmentation

The Czech segmenter follows all of the general segmentation rules in the whitespace languages.

4.1.1.3.5 Danish Word Segmentation

The Danish segmenter follows all of the general rules for word segmentation for the whitespace languages.

The Danish segmenter keeps together plurals and possessives spelled with **s** or **'s**. Hyphens are not separated from compound parts written with a hyphen. Periods are not separated from ordinal digit expressions. Examples are shown below:

Text	Segmented
Eriks	Eriks
14.	14.
post- og telegrafvæsenet	post-
	og
	telegrafvæsenet

4.1.1.3.6 Dutch Word Segmentation

The Dutch segmenter follows all of the general segmentation rules in the whitespace languages and has some language-specific behavior.

The Dutch segmenter does not split contractions. Plurals and possessives spelled with **s** or **'s** are not split. Hyphens are not separated from compound parts written with a hyphen.

Text	Segmented
m'n	m'n
'k	'k
auto's	auto's
Jansens	Jansens
honden- en kattenvoer	honden- en kattenvoer

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.1.3.7 English Word Segmentation

The English segmenter follows all of the general segmentation rules in the whitespace languages, but it has certain language-specific behavior.

In English, contractions like **don't**, **can't** and **won't** are separated into their constituent syntactic units. **Ain't** is not separated, since there is no clearly correct way to break it. The possessive endings **'s** and **'** are separated from the words they modify.

Text	Segmented
can't	ca n't
won't	wo n't
it's	it 's
ain't	ain't

Text	Segmented
helper's	helper 's
helpers'	helpers '

Abbreviations are not split from their punctuation, but do get split from following hyphens. Hyphens that occur in between two abbreviations will not break the syntactic unit. Abbreviations are listed in a system dictionary as well as in a set of rules allowing for uppercase and lowercase letters as well as periods and optional hyphens.

Combinations of alphabetic, numeric, and optionally, punctuation characters are usually kept together. For example:

Text	Segmented
Apr.-	Apr. -
D-Nebr.	D-Nebr.
3a.m.	3a.m.
11Jan.	11Jan.
Mon.-Thurs.	Mon.-Thurs.
Bloomberg-U.S.	Bloomberg-U.S.

4.1.1.3.8 Farsi Word Segmentation

The Farsi segmenter follows all of the general segmentation rules in the whitespace languages, but it has the following language-specific behavior. Certain control characters are also used to mark word boundaries in addition to whitespace and punctuation. For instance, the "zero-width nonjoiner", which indicates a final form character, is used to delimit word boundaries even when no whitespace is present between the two syntactic units.

Whitespace

Whitespace is not used consistently in Farsi text following a word or punctuation mark. The Farsi segmenter breaks text before and after each punctuation mark, as shown below:

Text	Segmented
فروختند، آنها	فروختند
	،
	آنها

Final Forms

In Farsi, the final form of letters is used to indicate the end of the word. Final form character markers, such as the 'zero-width joiner' in Unicode and the 'no width non-break' in Windows, are treated like whitespace and are used to segment words. For example:

Text	Segmented
برگ درخت ها	برگ
	درخت
	ها

However, when a word ends in one of the characters without a final form (such as `ورژ د ذ ا`), and there is no whitespace separating it from the following word, then the segmenter does not recognize and segment the word. For instance, `افکارمردم` and `اخباربرگزیده`.

ASCII and Arabic Characters

Some Farsi texts may contain punctuation, numbers and words in ASCII characters, and sometimes Arabic Unicode is used instead of Farsi. The Farsi segmenter is able to recognize and segment ASCII and Arabic Unicode characters.

Multiword Units

The Farsi segmenter treats multiword units as a single unit. For instance, `آسمان خرا، گفت و گو`. Multiword units may be written with or without whitespace, and the segmenter handles both formats, as shown below:

Text	Segmented
زد و خورد	زد و خورد
زدو خورد	زدو خورد

Hyphens are not included in Farsi multiword units or abbreviation analysis.

Punctuation

Apart from sentential punctuation (such as period, exclamation point, and question mark), all other punctuation is split from the preceding and/or following word. This is because whitespace between syntactic units or words is optional in Farsi, and is sometimes dropped from text. Also, words like **Yahoo!** do not exist, therefore the exclamation is not kept with the syntactic units.

Parentheses

Parentheses are split apart regardless of the syntactic unit that appears within the parentheses. There are no such cases in Farsi similar to **fox(es)** or **(re)analyze** in English.

Related Information

[English Word Segmentation \[page 17\]](#)

4.1.1.3.9 French Word Segmentation

The French segmenter follows all of the general segmentation rules in the whitespace languages.

French clitics and elisions are separated from the words they modify. The segmenter strips the hyphen from the verb form and keeps it with the following clitic. When separating elisions, the apostrophe is kept with the word whose letters were elided. Abbreviations are kept together with their punctuation.

Text	Segmented
donne-le-moi	donne
	-le
	-moi
l'abri	l'
	abri
trad.	trad.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.10 German Word Segmentation

The German segmenter follows all of the general segmentation rules in the whitespace languages.

The German segmenter splits contractions at the apostrophe. A few non-contractions that include apostrophes are not split at the apostrophe, because the apostrophe is part of the word.

Text	Segmented
geht's	geht
	's
auf's	auf
	's
Maxime's	Maxime's

When a compound consists of two parts joined by a conjunction, the hyphen is not separated from the partial compound. Leading hyphens are not split off if the following word begins with a lowercase letter. However, if the following word begins with an uppercase letter, the hyphen is split off.

Text	Segmented
West- und Ostgoten	West-
	und
	Ostgoten
Silbermesser und -gabel	Silbermesser
	und
	-gabel
-West	-
	West

Abbreviations are not split off from their punctuation. Ordinal numbers are also kept together with their period.

Text	Segmented
Mrd.	Mrd.

Text	Segmented
bzgl.	bzgl.
43.	43.

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.1.3.11 Greek Word Segmentation

The Greek segmenter follows all of the general segmentation rules in the whitespace languages.

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.1.3.12 Hebrew Word Segmentation

The Hebrew segmenter follows the general segmentation rules for whitespace languages. The Hebrew segmenter has certain language-specific behavior with respect to quotation marks and apostrophes.

A quotation mark " is used as an abbreviation marker in Hebrew, and when found in penultimate position, the " is not separated from the word as a quotation mark. Rather, it is treated as part of the word.

Text	Segmented
ברא"ה	ברא"ה

The apostrophe ' is sometimes used to modify the sound of the characters **gimel**, **tsadi** and **zayin**. This is used in borrowed words where the foreign sound does not match the pronunciation of the Hebrew characters. For example, when the apostrophe occurs to the left of **gimel**, its sound is changed from the **g** in good to the **j** sound in **George**. In these cases, the apostrophe is segmented together with its word, as shown below.

	Text	Segmented
	ג'ורג	ג'ורג

4.1.1.3.13 Hungarian Word Segmentation

The Hungarian segmenter follows all of the general segmentation rules in the whitespace languages.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.14 Indonesian Word Segmentation

The Indonesian segmenter follows all of the general segmentation rules in the whitespace languages. There are no apostrophe contractions in Indonesian.

4.1.1.3.15 Italian Word Segmentation

The Italian segmenter follows all of the general segmentation rules in the whitespace languages.

The segmenter separates Italian elisions, including elided numbers, from the words they modify. When separating elisions, the apostrophe is kept with the word whose letters were elided. Combined words that are written without apostrophes are not split.

Text	Segmented
d'un'artistica	d' un' artistica
cinqu'invitati	cinqu' invitati
nella	nella

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.16 Korean Word Segmentation

The Korean segmenter uses the same underlying algorithm as the segmenters for the whitespace languages. In brief, punctuation characters are treated as syntactic unit delimiters, as are the whitespace and tab characters.

Korean words can contain several sorts of dependent morphemes, such as case markers and inflectional endings. The dependent morphemes do not become separate words.

For example, noun **사람들은** "person-PL-TOP" is segmented as one word even though it consists of three morphemes, the noun **사람**, the plural marker **들**, and the topic marker **은**. Similarly, in **가셨습니다** ("a (respectable person) has gone"), the subject honorific **시**, the past tense suffix **었**, which contract together as **셨** the addressee honorific suffix **습**, the indicative suffix **니**, and the declarative suffix **다** occur in that order after the head verb stem **가** ("go").

Text	Segmented
그	그
사람들은	사람들은
못됐다	못됐다
.	.

Text	Segmented
선생님께서는	선생님께서는
벌써	벌써
서울로	서울로
가셨습니다	가셨습니다
.	.

Segmenters for European languages recognize multiword units as a single unit (for example, "**to and fro**" in English). The Korean segmenter gives the same treatment to phrases like **이랬다** and **저랬다**.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.17 Neutral Language Word Segmentation

The Neutral segmenter follows all of the general segmentation rules in the whitespace languages.

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.1.3.18 Norwegian (Bokmål) Word Segmentation

The Bokmål segmenter follows all of the general segmentation rules in the whitespace languages.

The Bokmål segmenter does not split plurals and possessives spelled with **s** or **'s**. Hyphens are not separated from compound parts written with a hyphen. Periods are not separated from ordinal digit expressions.

Text	Segmented
Eriks	Eriks
32.	32.
lonns- og inntektsutviklingen	lonns- og inntektsutviklingen

4.1.1.3.19 Norwegian (Nynorsk) Word Segmentation

The Nynorsk segmenter follows all of the general segmentation rules in the whitespace languages.

The Nynorsk segmenter does not split plurals and possessives spelled with **s** or **'s**. Hyphens are not separated from compound parts written with a hyphen. Periods are not separated from ordinal digit expressions.

Text	Segmented
Eriks	Eriks
32.	32.

Text	Segmented
lonns- og inntektsutviklinga	lonns-
	og
	inntektsutviklinga

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.1.3.20 Polish Word Segmentation

The Polish segmenter follows all of the general segmentation rules in the whitespace languages.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.21 Portuguese Word Segmentation

The Portuguese segmenter follows all of the general segmentation rules in the whitespace languages and has some language-specific behavior.

Clitics are not split off, and combined words are treated as one word.

Text	Segmented
dir-se-ia	dir-se-ia
pela	pela

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.22 Romanian Word Segmentation

The Romanian segmenter follows all of the general segmentation rules in the whitespace languages.

Romanian clitic pronouns and particles are separated from the words they modify. The segmenter strips the hyphen from the verb and keeps it with the following or preceding clitic:

Text	Segmented
rugându-mă	rugându -mă
n-a	n- a

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.23 Russian Word Segmentation

The Russian segmenter follows all of the general segmentation rules in the whitespace languages. The Russian segmenter handles multiword units such as **вряд ли**, **35-й**, **Европа Плюс**, as well as abbreviations like **лаб** and **фр**.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.24 Serbian (Cyrillic) Word Segmentation

The Serbian (Cyrillic characters) segmenter follows all of the general segmentation rules in the whitespace languages.

4.1.1.3.25 Serbian (Latin) Word Segmentation

The Serbian (Latin characters) segmenter follows all of the general segmentation rules in the whitespace languages.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.26 Slovak Word Segmentation

The Slovak segmenter follows all of the general segmentation rules in the whitespace languages.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.27 Slovenian Word Segmentation

The Slovenian segmenter follows all of the general segmentation rules in the whitespace languages.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.1.3.28 Spanish Word Segmentation

The Spanish segmenter follows all of the general segmentation rules in the whitespace languages, but it also has certain language-specific behavior.

Contracted words such as **del** or **al** are not split. Clitics are not separated, for example, in **dámelo**. Trailing hyphens are split apart from their words. Ordinal numbers are not separated from their period.

Text	Segmented
dámelo	dámelo
del	del
empresa-	empresa
	-
2a.	2a.

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.1.3.29 Swedish Word Segmentation

The Swedish segmenter follows all of the general segmentation rules in the whitespace languages.

The Swedish segmenter does not split plurals and possessives spelled with **s** or **'s**. Hyphens are not separated from compound parts written with a hyphen. Numeric and punctuation combinations are kept together.

Text	Segmented
Eriks	Eriks
metall- och kemikoncern	metall-
	och
	kemikoncern
456:-	456:-

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.1.3.30 Turkish Word Segmentation

The Turkish segmenter follows all of the general segmentation rules in the whitespace languages.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.2 Nonwhitespace Languages

Nonwhitespace languages do not separate words with whitespace.

Nonwhitespace languages include the Chinese languages, Japanese, and Thai (CCJT for short). Word segmentation in the CCJT languages occurs with a slightly different algorithm due to their structure. Because complete morphological analysis is required to perform word segmentation in these languages, the word segmentation, stemming, and part-of-speech tagging operations occur in a single step.

4.1.2.1 Language-Specific Nonwhitespace Segmentation Examples

A collection of examples of word segmentation for the nonwhitespace languages.

4.1.2.1.1 Simplified and Traditional Chinese Word Segmentation

The Chinese segmenter follows all of the general segmentation rules in the nonwhitespace languages, but it has the following language-specific behavior.

Bound morphemes like affixes are attached to content words. Also, classifiers are attached to preceding numbers. In the following example, 多 in 多媒体 is a prefix and 台 in 三台 is a classifier.

Text	Segmented
门市	门市
经营	经营
部门	部门
购得	购得

Text	Segmented
多媒体	多媒体
电脑	电脑
三台	三台

Hyphenated words are segmented into their separate parts. For instance:

Text	Segmented
北京 - 东京	北京
	-
	东京

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.2.1.1.1 Nongranular Word Segmentation in Simplified and Traditional Chinese

Because Chinese words are not inflected, the stems of all words are identical to their source forms.

This is true of the open class words listed in the following table as well as the closed class words.

Category	Base form	Example
Noun	Source form	政府 → 政府, 学生 → 学生
Verb	Source form	负责 → 负责, 保留 → 保留
Adjective	Source form	小 → 小, 必须 → 必须
Adverb	Source form	非常 → 非常

4.1.2.1.1.2 Granular Word Segmentation in Simplified and Traditional Chinese

The granular module output differs from the nongranular module in that classifiers are separated from numerals, prefixes and suffixes are separated from their head words, and compound analysis is performed.

The output of the modules is designed for optimized text indexing and search systems.

Examples are shown below.

Classifiers are separated from numerals:

Text	Output
一本	—
	本

Prefixes and suffixes are separated from their head words:

Text	Output
女教师	女
	教师
小张	小
	张
发展部	发展
	部

Compounds are broken into their separate components:

Text	Output
布赖斯峡谷国家公园	布赖斯
	峡谷
	国家
	公园
彩色监定系统	彩色
	监定
	系统

The granular variant supports all the same operations as the nongranular Chinese modules. However, its fine-grained output provides less contextual information for each term, and this ambiguity can compromise the accuracy of the tagging operations. For these operations, we recommend using the nongranular Chinese module. The granular variant is recommended for stemming purposes only.

4.1.2.1.2 Japanese Word Segmentation

The Japanese segmenter follows all of the general segmentation rules in the nonwhitespace languages, but it has the following language-specific behavior.

The word segmenter breaks text into minimum syntactic units, called "bunsetu" in Japanese. A Japanese word, or syntactic unit, contains a number of dependent words such as case markers (for example, が, を, は) and inflectional parts of the predicate (for example, ます, た, させる).

Case markers are separated from nominal elements:

Text	Segmented
太郎が本を読む	太郎
	が
	本
	を
	読む

Inflectional suffixes are segmented together with the head verb:

Text	Segmented
来ました	来ました
食べさせられました	食べさせられました

Aspectual and modal verbs are separated from the head verb:

Text	Segmented
食べるようだ	食べる
	よう
	だ
食べはじめた	食べ
	はじめた

The copula is separated from the head noun:

Text	Segmented
本だ	本

Text	Segmented
	だ

Classifiers are attached to the preceding numeral:

Text	Segmented
3冊	3冊
三人	三人

Noun prefixes are separated from the nouns, while the verb and adjectival prefixes are attached to their heads:

Text	Segmented
高品質	高 品質
お座り	お座り
バカでかい	バカ でかい

Punctuation marks, including opening and closing marks, are segmented separately:

Text	Segmented
「紅花」	「 紅花 」
★注意	★ 注意

The Japanese segmenters treat whitespace within hiragana and katakana sequences as syntactic unit boundaries. That is, syntactic units are broken as follows:

Text	Segmented
オフィスソリューション	オフィス ソリューション

Newline characters, such as "/n", are preserved when surrounded by katakana words.

Hyphens and slashes also break syntactic units. Hyphenated kanji words and katakana words are separated:

Text	Segmented
東京—箱根	東京 — 箱根
パリ—ロンドン	パリ — ロンドン
パリ / ロンドン	パリ / ロンドン

Numeric expressions with or without punctuation marks are kept together:

Text	Segmented
12,000	12,000
20/20	20/20
25%	25%
2 . 5	2 . 5
二五—五十	二五—五十

i Note

Japanese words written entirely in hiragana script, rather than in the more standard combination of kanji and hiragana, may not be properly segmented due to ambiguity. Such written style is usually restricted to text intended for children or learners of Japanese.

Related Information

[Word Segmentation \[page 14\]](#)

[Whitespace Languages \[page 14\]](#)

4.1.2.1.2.1 Nongranular Word Segmentation in Japanese

The Japanese nongranular segmenter follows the general segmentation rules.

In brief, the major word classes, also known as the open classes, stem to their base or dictionary forms. This is shown in the table below.

Category	Base form
Noun	Source form
Verb	Non-past-tense form
Adjective	Non-past-tense form
Adverb	Source form

Classifiers and derivational suffixes are not removed from nouns in the segmenter. For example,

Part of Speech	Word	Stem
Num + Cl	二冊	二冊
Nn + Adj_suffix	高さ	高い
Verb + Nn_suffix	読み方	読み方
Nn + Pl_suffix	学生たち	学生たち
Nn + Hon_suffix	佐藤様	佐藤様

Closed class words like pronouns, demonstratives, letters and numbers are segmented to their base form, and any existing case markers are dropped.

Japanese verb and adjective words can be inflected for tense, aspect, modality, politeness, and so on. The segmenter returns verbs and adjectives without any inflectional endings. This is commonly called the dictionary form. For example, the following inflected verb forms all stem to 食べる ("eat").

Word	Stem
食べた	食べる
食べさせる	食べる
食べます	食べる
食べない	食べる

4.1.2.1.2.2 Granular Word Segmentation in Japanese

The granular Japanese language module provides more fine-grained segmentation and stemming results than the nongranular module. Its output is designed for optimized text indexing and search systems.

The granular module output differs from the nongranular module in that classifiers, numerals, prefixes and suffixes are separated from their head words, and compound analysis is performed.

Examples are shown below.

Classifiers are separated from numerals:

Text	Output
1996 年	1996
	年
30 分	30
	分

Prefixes are separated from their head words:

Text	Output
お部屋	お
	部屋
副作用	副
	作用

Suffixes are separated from their head words:

Text	Output
全国的	全国
	的
須田さん	須田
	さん
ニューヨーク州	ニューヨーク
	州

Compounds are broken into their separate components:

Text	Output
朝日新聞社	朝日 新聞 社
日本電信電話	日本 電信 電話
サウンドマスター	サウンド マスター

The granular variant supports all the same operations as the nongranular Japanese module. However, its fine-grained output provides less contextual information for each term, and this ambiguity can compromise the accuracy of the tagging operations. For these operations, we recommend using the nongranular Japanese module. The granular variant is recommended for stemming purposes only.

4.1.2.1.3 Thai Word Segmentation

The Thai segmenter follows all of the general segmentation rules in the nonwhitespace languages.

Suffixes and prefixes remain attached to their content words, as shown below:

Text	Segmented
ผู้สอน	ผู้สอน
โดยตั้งใจ	โดยตั้งใจ

i Note

Unlike Chinese and Japanese, Thai has only one segmentation mode at this time: nongranular.

Related Information

[Word Segmentation \[page 14\]](#)

4.1.2.1.3.1 Nongranular Word Segmentation in Thai

Because Thai words are not inflected, the stems of all words are identical to their source forms.

This is true of the open class words listed in the following table as well as the closed class words.

Category	Base form	Examples
Noun	Source form	ภาพถ่าย → ภาพถ่าย
Verb	Source form	กลับ → กลับ
Adjective	Source form	คอนกรีต → คอนกรีต
Adverb	Source form	กรวด → กรวด

4.2 Case Normalization

Case normalization provides case-normalized alternatives for words which, by their position in a sentence or because they occur in a title, may or may not appear with their inherent, meaningful capitalization.

For instance, a proper noun like SAP is always capitalized, but a common noun like horse is only capitalized if it begins a sentence or occurs in a title. Therefore, if Horse is encountered, the case normalizer provides the lower-case alternative so that later processing will not mistake Horse for a proper noun. The two resulting alternatives can then be passed on to the stemming or tagging operations.

i Note

Case normalization is not relevant to languages that do not distinguish between upper and lower case, for example, the CCJT languages, Arabic, Korean, Farsi, and Hebrew.

Case normalization depends on the type of sentence (normal sentence, title, or query) and the position of the word to be normalized in each sentence type. The important position to consider is the sentence-initial position, where special normalization rules may apply. Words directly following certain punctuation marks are also treated as if they are in sentence-initial position.

- Title sentence
All capitalized words are normalized. For example, a newspaper heading would be normalized as:
 - Cardinals Strike Out (Cardinals | cardinals) (Strike | strike) (Out | out)
- Query sentence
Lowercase words are normalized to their upper case variants. Capitalized and all-caps words are not normalized in query sentences.
 - aaaa: aaaa, Aaaa, AAAA
 - aaaA: aaaA, AaaA
- Normal sentence
Capitalized words are normalized when they occur in sentence-initial position. All-caps words in sentence-initial position are also normalized. In other positions of normal sentences, capitalized and all-uppercase words are not normalized. For instance:

- Aaaa bbb Cccc: (Aaaa | aaaa) (bbb) (Cccc)
- AAAA bbb CCCC: (AAAA | Aaaa | aaaa) (bbb) (CCCC)

If an unknown word is encountered, that is if none of the case-normalized variants is found, all these variants are returned. For example, suppose the input sentence begins with the unknown word `Fbzzz`. The case normalizer then returns both `Fbzzz` and `fbzzz`.

4.3 Stemming

Stemming a word means identifying its stem—its base form, the form referenced in a dictionary.

Words like **speaks** or **speaking** have one stem—**speak**. Some words have more than one possible stem: **spoke**, for instance, may turn out, in context, to be the past tense of the verb **speak**, but it could also be the singular form of the noun **spoke**. A stem is a base form for one or more variant (inflected or derived) forms found in text. The stem is typically the form used as the citation form in a dictionary entry.

Stemming a word means finding and returning its stem. For example, rather than redundantly deal with **grind**, **grinds**, **grinding**, **ground**, and so on, all of these source forms can be recognized as variants of the single verb **grind**. **Ground** can also be a noun whose meaning is completely unrelated to the verb **grind**.

The example of indexing documents according to key words they contain can help to better understand the advantages of working with more abstract forms. If indexing is done naïvely, **grind**, **grinds**, **grinding**, **ground** will be handled as unrelated words, and a query containing one of these variants will not return documents containing the other variants. With the use of a stemmer, however, all of the variants will be indexed under the base form **grind** (verb).

The stemmer the software uses receives input of a series of orthographic units (for example, **ground**) and associates each unit with one or more base forms (for example, **ground**, **grind**). The stemmer always returns all possible alternative stems for each input term.

Depending on the language, the stemmers can support different variants of the stemming operation:

- For some languages, stemmers return all possible normalized stems for the input.

Example	Stems to
{aller, vais, vas, va, allons, allez, vont} [French]	aller
{reach, reaches, reached, reaching}	reach
{big, bigger, biggest}	big
{balloon, balloons}	balloon
{go, goes, going, gone, went}	go

The **emphasized** words are the stems (dictionary forms). The characters added to the stem (**es** in **reaches**, **s** in **balloons**) are called inflections or affixes.

To handle unknown words such as neologisms, stemmers contain sets of morphological rules that apply to words.

- For other languages, stemmers cover the normalized stems and they allow wider variation in capitalization, accentuation, and similar features, as found in informal text such as e-mail and text messages.

Related Information

[Language-Specific Stemming Examples \[page 41\]](#)

4.3.1 Language-Specific Stemming Examples

4.3.1.1 Arabic Stemming

The major word classes in Arabic stem to the unvocalized stem form, rather than the root form used in traditional Arabic lexicography. This is because too much information is lost when stemming directly to the root. This is shown in the table below. In general, function words stem to themselves. Known proper nouns stem to themselves, and unknown proper nouns are returned unchanged.

Category	Base form	Examples
Noun	Singular surface form for proper nouns	جهة ← الجهات عمرو ← عمرو
Verb	Verb stem	أمل ← تأمل
Adjective	Masculine singular	الكبيرة ← الكبيرة
Adverb	Adverb stem	هنا ← هنا

4.3.1.2 Catalan Stemming

The Catalan stemmer follows the general stemming rules.

In brief, the major word classes, also known as the open classes, stem to their base forms. Proper nouns stem to themselves. Diminutive and superlative endings are removed from open classes. This is shown in the table below:

Category	Base form	Examples
Noun	Non-diminutive masculine singular	vedelleta → vedell
Proper Noun	Stemmed to themselves	Jordi → Jordi
Verb	Infinitive	considerava → considerar
Adverb	Source form	bé → bé, activament → activament
Pronoun	Masculine, nominative form	me → jo

Catalan pronouns are stemmed in the following way. All uninflected forms stem to themselves. All personal pronouns maintain their number information. If applicable, these pronouns are stemmed to the nominative form. All other forms stem to the masculine, singular form. This is shown in the table below:

Text	Stem
tothom	tothom
elles	ells
em	jo
aquestes	aquest

Closed class words like determiners and ordinal numbers are stemmed to the masculine, singular form. Noninflecting word categories stem to themselves, for example, conjunctions, cardinal numbers and prepositions:

Text	Stem
mitges	mig
ni	ni

Acronyms, abbreviations and multiword syntactic units stem to themselves:

Text	Stem
IVA	IVA
tel.	tel.
davant de	davant de

Contracted forms are stemmed into their component part:

Text	Stem
pel	per=el
als	a=el

Additionally, the Catalan stemmer does not require correct capitalization and accentuation and allows required hyphens to be optional.

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
CATALUNYA	Catalunya
catalunya	Catalunya

Diacritics

Characters without diacritics are allowed in place of characters with diacritics. Also, acute accents are allowed in place of grave accents and vice versa.

Example	Output
artistic	artístic
artístic	artístic

Hyphenation

Hyphens in non-numeric expressions are optional.

Example	Output
mezzo soprano	mezzo-soprano
mezzosoprano	mezzo-soprano

Related Information

[Stemming \[page 40\]](#)

4.3.1.3 Croatian Stemming

The Croatian stemmer follows the general stemming rules.

In brief, the major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	vojske → vojska , ljudi → èovjek , mjesta → mjesto
Verb	udovoljava → udovoljavati , ponude → ponuditi , komentirao → komentirati
Adjective	srbijansku → srbijanski , spremni → spreman , izborni → izboran
Adverb	kako → kako , sada → sada , opet → opet

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
zagreb	Zagreb
Zagreb	Zagreb

Diacritics

Characters without diacritics are allowed in place of characters with diacritics.

Example	Output
zagrebackom	zagrebački

Example	Output
vecina	večina

Hyphenation

Hyphens in non-numeric expressions are optional.

Example	Output
befitness-centar	fitness-centar
fitnesscentar	fitness-centar

Related Information

[Stemming \[page 40\]](#)

4.3.1.4 Czech Stemming

In brief, the major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	tělem → tělo, města → město, výzkumy → výzkum
Verb	máš → mít, vrátil → vrátit, dostane → dostat, pracuji → pracovat
Adjective	velká → velký, starší → starý
Adverb	brzy → brzy, dnes → dnes

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
praha	Praha
Praha	Praha
ceskoslovensku	Československo

Diacritics

Characters without diacritics are allowed in place of characters with diacritics.

Example	Output
americky	americký
dalsi	další

Related Information

[Stemming \[page 40\]](#)

4.3.1.5 Danish Stemming

The Danish stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

The base forms for Danish shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Base form	Example
Noun	Indefinite singular	kager → kage, cyklen → cykel
Verb	Infinitive	sendes → sende, luk → lukke
Adjective	Base form	kolde → kold, smukkeste → smuk
Adverb	Base form or source form	oftest → ofte, bagfra → bagfra

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
Danmark	Danmark
danmark	Danmark

Special Characters

In cases where Latin-1 characters are used to mimic special Danish characters, words are stemmed to their dictionary forms.

Example	Output
europæisk	europæisk
europaeisk	europæisk

Hyphenation

The stemmer accepts spelling variants even when obligatory hyphens are missing.

Example	Output
finsk-ugrisk	finsk-ugrisk
finskugrisk	finsk-ugrisk

Related Information

[Stemming \[page 40\]](#)

4.3.1.6 Dutch Stemming

The Dutch stemmer follows the general stemming rules. It addresses nouns, verbs, adjectives, and adverbs.

The major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Base form	Examples
Noun	Non-diminutive singular	bloem → bloem, emmers → emmer, kinderen → kind
Verb	Infinitive	schrijft → schrijven, hebt → hebben
Adjective	Base form	lange → lang, onhandigste → handig
Adverb	Base form or source form	eventjes → even, liefst → graag, gisteren → gisteren

Additionally, the Dutch stemmer does not require correct capitalization and accentuation and allows required hyphens to be optional. It also accepts certain nonstandard conventions for hyphens in compounds.

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
USA	USA
usa	USA

Diacritics

Characters without diacritics are allowed in place of characters with diacritics. Also, acute accents are allowed in place of grave accents and vice versa.

Example	Output
privé	privé
prive	privé
privè	privé

Hyphenation

Hyphens in non-numeric expressions are optional. The stemmer accepts these compounds even when obligatory hyphens are missing.

Example	Output
auto-ongeluk	auto#ongeluk
autoongeluk	auto#ongeluk
kinderbioscoop	kind#bioscoop
kinder-bioscoop	kind#bioscoop

Related Information

[Stemming \[page 40\]](#)

4.3.1.7 English Stemming

The English stemmer follows the general stemming rules. It addresses nouns, pronouns, verbs, adjectives, and adverbs. It also stems British spellings to American spellings.

In brief, the major word classes, also known as the open classes, stem to their base forms. This is shown in the table below.

Category	Base form	Examples
Noun	Singular	dog, dogs → dog
Verb	Infinitive	runs, ran, run → run
Adjective	Base form	happy, happier, happiest → happy
Adverb	Base form or source form	quickly → quickly

English pronouns are stemmed in the following way. All uninflecting forms stem to themselves. Plural-only forms and all personal pronouns maintain their number and gender information. If applicable, these pronouns are stemmed to the nominative form. All other forms stem to the singular form. This is shown in the table below:

Text	Stem
none	none
that	that
themselves	themselves
her	she
these	this

The stemmer handles the spelling variation found in American and British English. Both variants stem to the American spelling. These behaviors are shown in the following table:

Text	Stem
color	color
colour	color
organization	organization
organisation	organization

Additionally, the English stemmer does not require correct capitalization and accentuation and allows required hyphens to be optional.

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
Eric	Eric
eric	Eric

Diacritics

Characters without diacritics are allowed in place of characters with diacritics. Also, acute accents are allowed in place of grave accents and vice versa.

Example	Output
Champs Elysees	Champs Élysées
Champs Élysees	Champs Élysées
Champs Elysées	Champs Élysées
Champs Èlysees	Champs Élysées

Hyphenation

Hyphens in non-numeric expressions are optional.

Example	Output
square-dance	square-dance
squaredance	square-dance
motherinlaw	mother-in-law

Related Information

[Stemming \[page 40\]](#)

4.3.1.8 Farsi Stemming

The Farsi stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed. Word categories that do not inflect are stemmed to themselves.

Category	Base form	Text	Stem
Noun	Singular	مسافرین	مسافر
Irregular Noun	Singular	قوانین	قانون
Verb	Infinitive	می خندیدند	خندیدن
Adjective	Singular, base form	کوچکترین	کوچک
Adverb	Singular, base form	بیشتر	بیشتر

In Farsi, some inflectional affixes are always attached to the stem, while others are always separated by a final form character or whitespace. Still other affixes may occur either way. The Farsi stemmer removes the affixes uniformly for these variants, as shown below:

Text	Stem
فلسطینہا	فلسطینی
فلسطینی ها	فلسطینی
فلسطینی ها	فلسطینی
میروند	رفتن
می روند	رفتن

Cliticized elements and others not belonging to the stem are removed by the stemmer, including clitic pronouns, copulas, and prepositions:

Text	Stem
کتابهایمان	کتاب
وکیلت	وکیل
بشیوه	شیوه
درمنزل	منزل
ازما	ما

The stemmer also removes affixes from deverbal nouns and adjectives and stems these forms to the verb's infinitive form:

Text	Stem
بینندگان	دیدن
بریده	بریدن

Caveats

In Farsi, some clitics and prefixed prepositions are optionally attached to the stem. The output of the stemmer differs depending on whether these elements are attached or not. For example, if the input is two syntactic units—a preposition followed by a noun, the stemmer outputs two stems. If the preposition is attached to the noun, then the stemmer removes it and returns the noun stem.

Text	Stem
به دفتر	به دفتر
بدفتر	دفتر

Complex verbal constructions are stemmed to their component parts rather than being stemmed as one verbal construction. For instance, both the past tense *کار کردند* and present tense forms *می کنند کار می* of the complex verb *کار کردن* 'work' are stemmed to the noun *کار* 'work' followed by the verb *کردن* 'do'.

If your document contains diacritics or stylistic markers such as short vowels, "tashdid," "sokun," or "keshideh" characters, the syntactic units may not be recognized by the stemmer module.

Note that there is no derivational component in Farsi (Persian) but there is limited derivational stemming in the regular stemmer module for deverbal nouns and adjectives (nouns and adjectives formed from verbs).

Abbreviations

Abbreviations that end in a period and do not have an intervening space are recognized and tagged as "Abbreviation" by the word-breaking process. For example:

م.ه.ق.

Abbreviations with a space or without a period are treated in the stemmer. For example:

ه ق

ق.م

Acronyms are also treated in the stemmer. For example:

بی بی سی

4.3.1.9 French Stemming

The French stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms. Feminine occupational nouns stem to their masculine counterparts. Proper nouns stem to themselves.

This stemming is shown in the table below.

Category	Base form	Examples
Noun	Singular	livres → livre; actrice → acteur
Proper Noun	Source form	France → France
Verb	Infinitive	connais, connaissez → connaître
Adjective	Masculine singular	grandes, grande → grand
Adverb	Source form	probablement → probablement

French pronouns are stemmed in the following way. All uninflecting forms stem to themselves. Plural-only forms and all personal pronouns maintain their number information. If applicable, these pronouns are stemmed to the nominative form. All other forms stem to the masculine, singular form. This is shown in the table below:

Text	Stem
beaucoup	beaucoup
plusieurs	plusieurs
elles	ils
moi	je
lesquelles	lequel

Closed class words may be regularized or they may stem to themselves.

These word categories stem to themselves: abbreviations, acronyms, interjections, numbers and onomatopoeia forms. This is shown in the table below:

Example	Stem
par ex.	par ex.
min.	min.
UNICEF	UNICEF

Contracted prepositions are broken into their component parts, and these stems are returned with an equal sign in between, indicating that the stems are of equal semantic importance. If the contracted preposition occurs in a multiword units, then the final contraction is broken. This is shown in the following table:

Example	Stem
au	à=le
au moment du	au moment de=le

Additionally, the French stemmer does not require correct capitalization and accentuation and allows required hyphens to be optional.

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
Eric	Eric
eric	Eric

Diacritics

Characters without diacritics are allowed in place of characters with diacritics.

Example	Output
Champs-Elysees	Champs-Élysées
Champs-Élysees	Champs-Élysées
Champs-Elysées	Champs-Élysées

Hyphenation

Hyphens in non-numeric expressions are optional.

Example	Output
belle-mère	belle-mère
bellemère	belle-mère
lavevaisselle	lave-vaisselle

Related Information

[Stemming \[page 40\]](#)

4.3.1.10 German Stemming

The German stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

Basic German stemming is shown in the table below.

Category	Base form	Examples
Noun	Nominative singular	Tischen → Tisch; Leuten → Leute
Verb	Infinitive	schwimmt, schwamm, geschwommen → schwimmen

Category	Base form	Examples
Adjective	Base form	farbigen → farbig; vag → vage
Adverb	Source form	ganztags → ganztags

German pronouns are stemmed in the following way. All uninflecting forms stem to themselves. Plural-only forms and all personal pronouns maintain their number and gender information. If applicable, these pronouns are stemmed to the nominative form. All other forms stem to the singular, nominative form of the given gender (if applicable). This is shown in the table below:

Text	Stem
manch	manch
ich, meiner, mir	ich
demjenigen	dasjenige,derjenige

Uninflecting categories stem to themselves, for example, abbreviations, acronyms, numbers, conjunctions, and so on. The German stemmers support various spelling variants and normalize them all to a single form. The following table shows some examples:

Text	Stem
zahlr.	zahlr.
ZDF	ZDF
Delphin, Delfin	Delfin
behende, behände	behände

Contracted prepositions are broken into their component parts, and these stems are returned with an equal sign in between, indicating that the stems are of equal semantic importance. This is shown in the following table:

Text	Stem
aufs	auf=das
beim	bei=das,bei=der
zur	zu=die

Additionally, the German stemmer does not require correct capitalization and accentuation and allows required hyphens to be optional.

Case Variants

The stemmer accepts lower case letters in addition to capital letters for those words where the capitals are obligatory.

Example	Output
USA	USA
usa	USA

Hyphenation

Hyphens in certain German words are optional. Different variants are stemmed to a single standard form."

Example	Output
Achlaut	Achlaut
Ach-laut	Achlaut

Related Information

[Stemming \[page 40\]](#)

4.3.1.11 Greek Stemming

The Greek stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	αποζημιώσεις → αποζημίωση, όζαινες → όζαινα, ευμάρειαν → ευμάρεια
Verb	κοροϊδεμένη → κοροϊδεύω, φουκτωθούμε → φουκτώνω, μισθοδοτημένους → μισθοδοτώ
Adjective	αρθρωτήν → αρθρωτός, πλαστικές → πλαστικός, μονόσπερμοι → μονόσπερμος

Category	Examples
Adverb	πόθεν → πόθεν, κατανυκτικότερα → κατανυκτικά

Related Information

[Stemming \[page 40\]](#)

4.3.1.12 Hebrew Stemming

The Hebrew stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

Category	Base form	Examples
Noun	singular	פגישות-פגישה לוחות-לוח שירים-שיר
Verb	past, 3rd person, masculine, singular form	מבינים-הבין סיפר-ספר
Adjective	masculine, singular form	צעירות-צעיר טובה-טוב קטנה-קטן
Adverb	surface form	אף-אף אפילו-אפילו

Clitics are removed from their head words, as shown here:

Text	Stem
בשוק	שוק
העתונאים	עתונאי
עצין	עץ
מנורותיכן	מנורה

4.3.1.13 Hungarian Stemming

The Hungarian stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	hírt → hír, vállalatnak → vállalat, bevételei → bevétel
Verb	kérek → kér, ünnepelnek → ünnepel, élünk → él
Adjective	privatizációssal → privatizációs, frisset → friss, japánok → japán
Adverb	már → már, majd → majd

Related Information

[Stemming \[page 40\]](#)

4.3.1.14 Indonesian Stemming

The Indonesian stemmer follows the general stemming rules, but it has certain language-specific behavior. The stemmer for Indonesian delivers the root word for an entry.

In identifying the root word, it can cross word categories when necessary. For example, the noun **permainan** is derived from the verb **main** by adding the circumfix **per-an**. The stemmer is designed to work for certain parts of speech in Indonesian, and is the default stemming behavior.

Verb lemmas are stemmed back when they have noun and verb derivations.

Text	Stem
permainan (Noun)	main (Verb)
penjual (Noun)	jual (Verb)
membeli (Verb)	beli (Verb)

Noun lemmas are stemmed back when they have adjective and verb derivations, but not for noun derivations.

Text	Stem
mengurbanisasikan (Verb)	urbanisasi (Noun)
kebaya-bayian (Adjective)	bayi (Noun)

Text	Stem
pembukuan (Noun)	pembukuan (Noun)

Adjective lemmas are stemmed back for adjective, adverb, and verb derivations.

Text	Stem
kekecilan (Adjective)	kecil (Adjective)
meninggikan (Verb)	tinggi (Adjective)
pelan-pelan (Adverb)	pelan (Adjective)
keikhlasan (Noun)	keikhlasan (Noun)

Basic adverb lemmas are stemmed back for adverb, and verb derivations.

Text	Stem
berpura-pura (Verb)	pura-pura (Adverb)
hampir-hampir (Adverb)	hampir (Adverb)

Plural reduplication form of noun lemmas, such as **buku-buku** stem back to their lemma **buku**. Proper nouns stem to themselves. Prefix, infix, and suffixes attached to verb lemmas stem back to the verb lemma themselves. Indonesian does not have tense or gender inflections.

Category	Base From,	Examples
Noun	Lemma	buku-buku → buku
Proper Noun	Source form	Jakarta → Jakarta
Verb	Lemma	mengirimkan → kirim, kupinjam → pinjam

Additionally, the Indonesian stemmer does not require correct capitalization and allows required hyphens to be optional.

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
INDONESIA	Indonesia
indonesia	Indonesia

Hyphenation

Hyphens in non-numeric expressions are optional.

Example	Output
foya foya	foya-foya
foyafoya	foya-foya

4.3.1.15 Italian Stemming

The Italian stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

This stemming is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Base form	Examples
Noun	Singular	capi → capo, pagine → pagina
Verb	Infinitive	andiamo → andare; parlava → parlare
Adjective	Masculine singular	alte → alto; grandissimo → grande
Adverb	Source form	felicemente → felicemente; più → più

Contracted prepositions and pronouns are broken into their component parts, and these stems are returned with an equal sign in between, indicating that the stems are of equal semantic importance. This is shown in the following table:

Example	Stem
allo	a=lo
glielo	lui=lui

Additionally, the Italian stemmer does not require correct capitalization and accentuation and allows required hyphens to be optional.

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
Roma	Roma
roma	Roma
USA	USA
usa	USA

Diacritics

Characters without diacritics are allowed in place of characters with diacritics. Also, acute accents are allowed in place of grave accents and vice versa.

Example	Output
città	città
citta	città
è	è, essere
é	è, essere

Hyphenation

Hyphens in non-numeric expressions are optional.

Example	Output
Clermont-Ferrand	Clermont-Ferrand
ClermontFerrand	Clermont-Ferrand
liberal-democratico	liberal-democratico
liberaldemocratico	liberal-democratico

Related Information

[Stemming \[page 40\]](#)

4.3.1.16 Korean Stemming

The Korean stemmer follows the general stemming rules.

In brief, the major word classes, also known as the open classes, stem to their base forms, or citation forms.

The stemmer for Korean does not break compound nouns.

Category	Base form
Noun/Pronoun	Base form (without case marking)
Verb	Declarative form
Adjective	Declarative form
Adverb	Source form
Numerical	Number form

A Korean noun word may include a postposition (particle) indicating case marking. The stemmer returns the normalized noun (such as the uninflected noun head or content word) minus any case marking.

Nouns, pronouns, proper nouns and numerals all stem to the base form, without any case markers. For example, the following nouns are all stemmed to 학생:

Word	Stem
학생이	학생
학생을	학생
학생까지	학생
학생한테서	학생
학생하고	학생
학생까지만	학생

Verbs and adjectives stem to the dictionary form, without any inflectional suffixes. The following inflected verb forms are all stemmed to 먹다 ("eat"):

Word	Stem
먹었다	먹다
먹었겠다	먹다

The sentence 학생이 케이크 마지막 조각을 먹었다. ("The student has eaten the last piece of cake.") is stemmed as follows:

Word	Stem
학생이	학생
케이크	케이크
마지막	마지막
조각을	조각
먹었다	먹다

4.3.1.17 Neutral Language Stemming

The Neutral language stemmer stems each word to itself.

4.3.1.18 Norwegian (Bokmål) Stemming

The Bokmål stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

This stemming is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Base form	Examples
Noun	Indefinite singular	dammer → dam; bondens → bonde
Verb	Infinitive	ventet → vente; sendes → sende
Adjective	Base form	laveste → lav; kalde → kald
Adverb	Base form or source form	nærest → nær; imens → imens

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
Norge	Norge

Example	Output
norge	Norge

Special Characters

In cases where Latin-1 characters are used to mimic special Norwegian characters, words are stemmed to their dictionary forms.

Example	Output
Tromsø	Tromsø
Tromso	Tromsø

Hyphenation

The stemmer accepts spelling variants even when obligatory hyphens are missing.

Example	Output
Nord-Korea	Nord-Korea
nordkorea	Nord-Korea

Related Information

[Stemming \[page 40\]](#)

4.3.1.19 Norwegian (Nynorsk) Stemming

The Nynorsk stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

This stemming is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Base form	Examples
Noun	Indefinite singular	bilen → bil; jenta → jente
Verb	Infinitive	leikte → leike; speil → speile
Adjective	Base form	høgare → høg; blått → blå
Adverb	Base form or source form	svintare → svint; imedan → imedan

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
Noreg	Noreg
noreg	Noreg

Special Characters

In cases where Latin-1 characters are used to mimic special Norwegian characters, words are stemmed to their dictionary forms.

Example	Output
Tromsø	Tromsø
Tromso	Tromsø

Hyphenation

The stemmer accepts spelling variants even when obligatory hyphens are missing.

Example	Output
ikkje-metallisk	ikkje-metallisk
ikkjemetallisk	ikkje-metallisk

Related Information

[Stemming \[page 40\]](#)

4.3.1.20 Polish Stemming

The Polish stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	hosannami → hosanna, fyrgolu → fyrgol
Verb	śle → stać, zajętego → zająć
Adjective	profonicznym → profoniczny, progościnniejsze → progościnnie
Adverb	procale → procały, wtyczkowie → wtyczkowy

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
warszawie	Warszawa
niemcu	Niemiec

Special Characters

Accented Polish characters are interchangeable with ASCII characters.

Example	Output
zrobic	zrobić
zandarm	żandarm

Example	Output
slucha	stuchać

Related Information

[Stemming \[page 40\]](#)

4.3.1.21 Portuguese Stemming

The Portuguese stemmer follows the general stemming rules.

The major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Base form	Examples
Noun	(Masculine) non-diminutive singular	filhos → filho, balinha → bala
Verb	Infinitive	traremos → trazer, alimentará → alimentar
Adjective	Masculine singular	bonitona → bonito, caríssimos → caro
Adverb	Positive form or source form	ultimamente → ultimamente, pessimamente → pessimamente, mal → mal

Contracted prepositions and pronouns are broken into their component parts, and these stems are returned with an equal sign in between, indicating that the stems are of equal semantic importance. If the contracted preposition occurs in a multiword unit, then the final contraction is broken. This is shown in the following table:

Example	Stem
pelo	por=o
dele	de=ele
abaixo deste	abaixo de=este
ma	eu=ela

Additionally, the Portuguese stemmer does not require correct capitalization and accentuation and allows required hyphens to be optional.

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
Rodrigo	Rodrigo
rodrigo	Rodrigo

Diacritics

Characters without diacritics are allowed in place of characters with diacritics.

Example	Output
acafrao	açafração
quilometro	quilômetro
pos-graduacao	pós-graduação

Hyphenation

Hyphens in non-numeric expressions are optional.

Example	Output
afro-asiático	afro-asiático
afroasiático	afro-asiático
pre-historia	pré-história
prehistoria	pré-história

4.3.1.22 Romanian Stemming

The Romanian stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	profesorul → profesor, muzeele → muzeu, marii → mare
Verb	terminam → termina, doresc → dori, credeam → crede
Adjective	frumoasa → frumos, mici → mic, eficace → eficace
Adverb	aici → aici, teoretic → teoretic, mai → mai

Additionally, the Romanian stemmer does not require correct capitalization and accentuation and allows required hyphens to be optional.

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
ROMÂNIA	România
românia	România

Special Characters

Characters without diacritics are allowed in place of characters with diacritics.

Example	Output
Romania	România
Stefan	Ștefan

Hyphenation

Hyphens in non-numeric expressions are optional.

Example	Output
week end	week-end

Example	Output
weekend	week-end

Related Information

[Stemming \[page 40\]](#)

4.3.1.23 Russian Stemming

The Russian stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

Basic Russian stemming is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	работу → работа, изменения → изменение
Verb	покупаю → покупать, едешь → ехать покупала → покупать, ехали → ехать
Adjective	красного → красный, краснее → красный, краснейшим → красный
Adverb	хорошо → хорошо, ясно → ясно

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
России	Россия
россии	Россия
Москве	Москва
москве	Москва

Special Characters

Russian letter “ё” is interchangeable with “е”.

Example	Output
тёмный	темный
темный	темный

Related Information

[Stemming \[page 40\]](#)

4.3.1.24 Serbian (Cyrillic) Stemming

The Serbian stemmer for the Cyrillic alphabet follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	детета → дете, борби → борба, земаља → земље
Verb	одржан → одржати, зна → знати, поведу → повести
Adjective	националистичке → националистички, бошњачким → бошњачки, озбиљним → озбиљан
Adverb	недовољно → недовољно, заједно → заједно

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
београд	Београд
Простор	простор

Related Information

[Stemming \[page 40\]](#)

4.3.1.25 Serbian (Latin) Stemming

The Serbian stemmer for the Latin alphabet follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	nemanja → nemanje, teglu → tegla, odgovorom → odgovor
Verb	donese → doneti, ponude → ponuditi, zadovoljimo → zadovoljiti
Adjective	srbijansku → srbijanski, spremni → spreman, izborni → izboran
Adverb	zato → zato, tako → tako, ubrzo → ubrzo, radikalno → radikalno

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
beograd	Beograd
Prostor	prostor

Special Characters

Characters with diacritics are interchangeable with ASCII characters.

Example	Output
coveka	čovek
utisati	utišati

Example	Output
mrsavost	mršavost

Related Information

[Stemming \[page 40\]](#)

4.3.1.26 Slovak Stemming

The Slovak stemmer follows the general stemming rules.

In brief, the major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	jablká → jablko, brány → brána, domom → dom, stoly → stôl
Verb	chcel → chciť, prosím → prosiť, boli → byť, myslí → myslieť
Adjective	tmavom → tmavý, úzkej → úzký, stará → starý
Adverb	dobre → dobre, nikde → nikde, neskôr → neskôr

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
bratislava	Bratislava
Bratislava	Bratislava

Diacritics

Characters without diacritics are allowed in place of characters with diacritics

Example	Output
odstupit	odstúpit
pat	pät

Related Information

[Stemming \[page 40\]](#)

4.3.1.27 Slovenian Stemming

The Slovenian stemmer follows the general stemming rules.

In brief, the major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	čtiva → čtivo, čtivu → čtivo, čbeličarju → čbeličar, čvrstost → čvrstost, čvrstostih → čvrstost
Verb	jva → jesti, jta → jesti, jte → jesti, je → jesti, jesla → jesti
Adjective	yorški → yorški, yorških → yorški, yorška → yorški
Adverb	čvrsto → čvrsto

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
ljubljana	Ljubljana
Ljubljana	Ljubljana

Diacritics

Characters without diacritics are allowed in place of characters with diacritics.

Example	Output
dolznost	dolžnost
demokraticen	demokratičen

Hyphenation

Hyphens in non-numeric expressions are optional.

Input	Output
grško-rimski	grško-rimski
grškorimski	grško-rimski

Related Information

[Stemming \[page 40\]](#)

4.3.1.28 Spanish Stemming

The Spanish stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms. Proper nouns stem to themselves. All diminutive endings are removed, even in given names.

Stemming is shown in the table below.

Category	Base form	Examples
Noun	Non-diminutive singular	caballitos → caballo; gatos → gato
Proper Noun	Non-diminutive source form	África → África; Anita → Ana
Verb	Infinitive	compuesto → componer; contéstame → contestar

Category	Base form	Examples
Adjective	Masculine singular	altas → alto; chiquito → chico
Adverb	Source form	por qué → por qué; cariñosamente → cariñosamente

Spanish pronouns are stemmed in the following way. All uninflecting forms stem to themselves. Plural-only forms and all personal pronouns maintain their number information. If applicable, these pronouns are stemmed to the nominative form. All other forms stem to the masculine, singular form. This is shown in the table below:

Text	Stem
algo	algo
ambas	ambos
ellas	ellos
mí	yo
éstas	éste

Closed class words like determiners and ordinal numbers are stemmed to the masculine, singular, nominative form. Non-inflecting word categories stem to themselves, for example, conjunctions, cardinal numbers and prepositions.

Text	Stem
esta	este
con	con

Acronyms, abbreviations and multiword units stem to themselves. Pronoun abbreviations stem to their full form. These behaviors are shown in the following table:

Text	Stem
UNAM	UNAM
p.ej.	p.ej.
Ud.	usted
los tuyos	el tuyo

Contracted words are stemmed into their component parts:

Text	Stem
conmigo	con=yo
al	a=el

Additionally, the Spanish stemmer does not require correct capitalization and accentuation and allows required hyphens to be optional.

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
Enrique	Enrique
enrique	Enrique

Diacritics

Characters without diacritics are allowed in place of characters with diacritics.

Example	Output
kilometro	kilómetro
kilómetro	kilómetro
compañia	compañía
compañia	compañía
compania	compañía
compañía	compañía

Hyphenation

Hyphens in non-numeric expressions are optional.

Example	Output
Alkaselzer	Alka-Selzer
Alka-Selzer	Alka-Selzer

4.3.1.29 Swedish Stemming

The Swedish stemmer follows the general stemming rules.

In brief, the major word classes, also known as the open classes, stem to their base forms. This is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Base form	Examples
Noun	Indefinite singular	hunden, hundar, hundarna → hund
Verb	Infinitive	springer, sprang, sprungit → springa
Adjective	Base form	vackra, vackert → vacker
Adverb	Base form or source form	snabbt → snabbt

Case Variants

Uppercase is allowed in place of lowercase and vice versa.

Example	Output
Sverige	Sverige
sverige	Sverige

Special Characters

In cases where Latin-1 characters are used to mimic special Swedish characters, words are stemmed to their dictionary forms.

Example	Output
Göteborg	Göteborg

Example	Output
Goteborg	Göteborg

Hyphenation

The stemmer accepts spelling variants even when obligatory hyphens are missing.

Example	Output
ekonomisk-politisk	ekonomisk-politisk
ekonomiskpolitisk	ekonomisk-politisk

Related Information

[Stemming \[page 40\]](#)

4.3.1.30 Turkish Stemming

The Turkish stemmer follows the general stemming rules. In brief, the major word classes, also known as the open classes, stem to their base forms.

This basic stemming is shown in the table below. Closed class words like determiners and pronouns may also be stemmed.

Category	Examples
Noun	kitaplar → kitap, anahtarını → anahtar, denizi → deniz
Verb	açabilir → aç, gideceğim → git, boyamaya → boya
Adjective	küçük → küçük, dikkatli → dikkat, güzel → güzel
Adverb	sessiz → ses, gergin → ger, çabukça → çabuk

Related Information

[Stemming \[page 40\]](#)

4.3.2 Stemming Unknown Words

The stemmer identifies the stems of all the standard words of a language. However, an unknown word, such as one not found in the system dictionary, will not have a stem. In general, the stemmer returns the input term as the stem itself.

Words that are unknown to the stemmer are handled by the stemmer guesser. The guesser contains morphological rules that predict how these unknown words should be stemmed.

The software provides stemmer guessers for Croatian, Czech, Dutch, English, French, German, Italian, Korean, Russian, Serbian, Slovak, Slovenian and Spanish.

4.4 Compound Analysis

Identifies and separates compound words, like **bookmark** or **birdbath**, formed by combining or concatenating several words. The software performs compound analysis for Dutch, German, Korean, and Swedish. The compounds are always separated into their components and the stems of the components are returned.

4.4.1 Language-Specific Compound Analysis Examples

Four languages offer compound analysis: Dutch, German, Korean and Swedish.

4.4.1.1 Dutch Compound Analysis

Compounding in Dutch can combine various parts of speech: nouns can combine with nouns, nouns with adjectives, and verbs with nouns.

Compounds are separated into their component stems. In the sample output below, the internal compound boundaries are marked by #.

Noun-Noun

Dutch noun compounds often incorporate linking elements. For instance, when the words **leven** and **echt** combine to form a compound, the linking element **-s-** is added between them, giving **levensecht**. The two most frequent linking elements are **-s-** and **-en-**. (A third linking element, **-e-**, occurs in only a few irregular compounds.)

Example	Output
begrafenisstoet	begrafenis#stoet
mensenrechtenorganisaties	mens#recht#organisatie
levensecht	leven#echt

Nouns with vowel changes or other irregularities are also handled. For example:

Example	Output
scheepskapitein	schip#kapitein
zonnenscherm	zon#scherm

If the first compound element ends in a vowel and the second element starts with a vowel, a hyphen is conventionally inserted between the elements in the source form. This hyphen does not appear in the stemmer output.

Example	Output
auto-ongeluk	auto#ongeluk

Compounds like *boeken- en platenzaak* ("book and record shop") or *kindertheater en -bioscoop* ("children's theater and cinema") are sometimes seen.

Example	Output
boeken- en platenzaak	boek en platenzaak
kindertheater en -bioscoop	kind#theater en bioscoop

Verb-Noun

Compounds can also combine verbs and nouns. The verb part is stemmed to the infinitive form of the verb.

Example	Output
schrijfwijze	schrijven#wijze

Noun-Adjective

In compounds combining adjectives and nouns, the linking elements often seen in noun-noun compounds are absent.

Example	Output
kinderloos	kind#loos

4.4.1.2 German Compound Analysis

The German stemming module includes a comprehensive mechanism for analyzing productive compounding, capable of handling an essentially infinite number of compound words.

Compounding in German can combine various parts of speech: nouns with nouns, nouns with adjectives, verbs with nouns, and so on. Hyphenated compounds are treated like other compounds.

By default, compounds are separated into their component stems.

In the sample output below, the internal compound boundaries are marked by #.

Example	Output
Muttertag	Mutter#Tag

The base form of a compound element is capitalized as it would be if it stood alone, no matter where it appears in the compound. Thus noun elements have capitalized stems, even if they are not the first element in the compound. Similarly, the stem of an adjective as the first element of a compound would not be capitalized.

Compounds like **Vor-** und **Nachmittag** ("before and after noon") or **Bachkonzerte und -kantaten** ("Bach concertos and cantatas") are sometimes encountered. The hyphen is not part of the base form.

Example	Output
Vor- und Nachmittag	vor
	und
	Nachmittag

Example	Output
Bachkonzerte und -kantaten	Bach#Konzert und Kantate

Noun-Noun Compounding

German noun-noun compounds often contain linking elements (**Fugenelemente**) between the main elements. For instance, when the words **Herr** ("gentleman") and **Mantel** ("coat") combine to form a compound, the linking element -en- is inserted between them, giving **Herrenmantel** (gentleman's coat). Linking elements do not appear in the stemmer output.

Example	Output
Herrenmantel	Herr#Mantel

The German module assumes that such linking elements can be determined based on the declension class of the noun which they follow.

Some examples of noun-noun compounds follow. Note that the module can also analyze compounds combining more than two words.

Example	Output
Lehrlingsnot	Lehrling#Not
Kinderarzt	Kind#Arzt
Kindesentführung	Kind#Entführung
Obstanbaugebiet	Obst#Anbau#Gebiet
Informatik-Konzepte	Informatik#Konzept

Nominal Final Elements

Nouns are not the only possible non-final elements of German compounds ending with nouns. The module also recognizes numerals, adjectives, adjectival participles, adverbs, verb stems, and proper nouns. As already noted, stem capitalization follows usage for single words.

Example	Output
Optimalsumme	optimal#Summe
Linkskurve	links#Kurve
Goethehaus	Goethe#Haus
Waschmaschine	waschen#Maschine

Adjectival Final Elements

An adjective can combine with other adjectives to form compound adjectives. Nouns and numerals can also be the first elements of compounds with adjectival final elements. For example:

Example	Output
blaugrün	blau#grün
hellgelb	hell#gelb
graphiktauglich	Graphik#tauglich
ausbaufähig	Ausbau#fähig
zweiblättrig	zwei#blättrig
blau-grau	blau#grau

4.4.1.3 Korean Compound Analysis

The Korean stemming module includes a comprehensive mechanism for analyzing productive compounding.

The sample output below shows the component stems as returned by the stemmer. The compound boundary is marked by #.

Noun-Verb Compound

A noun combines with an intransitive verb to create a compound verb. The noun is uninflected, and the verb can be inflected and will be stemmed to its base form.

Example	Output
계획이다	계획 # 이다
기술적인	기술적 # 이다
시간두고	시간 # 두다
해결되다	해결 # 되다

Noun-Noun Compounds

A noun combines with a noun to create another compound noun. The first noun is uninflected, and the second noun can be inflected and will be stemmed to its base form.

Example	Output
연기상을	연기 # 상
현대문제가	현대 # 문제
거래소시장에서	거래소 # 시장

Numerical Compounds

A numerical compound combines a number to create a complex numerical expression. The number is separated from its particle.

Example	Output
50만원씩	50만 # 원 # 씩
11시	11 # 시
4월부터	4 # 월 # 부터
5인치	5 # 인치
9%까지	9 # % # 까지
12월의	12 # 월 # 의

4.4.1.4 Swedish Compound Analysis

The Swedish stemming module includes a comprehensive mechanism for analyzing productive compounding. The sample output below shows the component stems as returned by the stemmer. The compound boundary is marked by #.

Noun-noun

Example	Output
aftonbladet	afton#blad
berglandskap	berg#landskap
flickskola	flicka#skola
kistryckel	kista#nyckel

Noun+s+noun

Example	Output
anfallsspel	anfall#spel
stavningsreform	stavnin#reform

4.5 Part-of-Speech Tagging

Identifying the grammatical category a word falls into, such as noun or verb, along with subclass attributes of each of these major categories, such as singular or plural for nouns, and present or past tense for verbs.

The part-of-speech tagger identifies and labels the part of speech for each word in context.

For certain of its language modules, the software supports the use of two types of part-of-speech tags. You can also use these tags when creating extraction rules:

Each supported language has its own set of part-of-speech tags, shown in the example topics.

- Umbrella tags—These tags identify major parts of speech at a high level, without breaking down the part of speech further than its overall function. For example, the `Nn` tag identifies all nouns, regardless of whether they are singular or plural, feminine or masculine, and so on.

- Complete tags—These tags identify the exact part of speech, along with its attributes. For example, the `Nn-P1` tag identifies plural nouns, and `V-Pres-3-sg` identifies present tense, 3rd person singular verbs.

Related Information

[Language-Specific Part-of-Speech Tagging Examples \[page 93\]](#)

4.5.1 Part-of-Speech Naming Conventions

Discusses the structure of tag names, and how to interpret them.

Tags consist of feature names separated by hyphens. The first feature name is called a category tag. It usually specifies the high level part of speech of the word, for example, noun or verb, abbreviated as `Nn` and `V` respectively. When the tag contains more than one part-of-speech, as in `V/Adj` or `Det/Pron`, this indicates that the part-of-speech can be of either category.

Feature tags classify the word more precisely. They may indicate number (for example, plural and singular), person (for example, first, second or third), or tense (for example, present and past). Thus, the tag `V-Pres-3-sg` indicates that the verb is present tense, third person singular.

When a feature appears in all lower case, as in the tag `Prep-para` from the Spanish tagger, it stands for a word in that language (here, Spanish `para`), and means that the word's distribution differs enough from that of other words of its category to rate its own feature. Such very specific features are listed in the language-specific tables.

4.5.2 Part-of-Speech and Stem Disambiguation

Provides complete linguistic analysis of input text, including selecting the “best choice” in context with respect to part-of-speech and stem information.

For example, in the input text “They ground coffee every morning”, the stem of “ground” is disambiguated based on the tag information:

Input	Part of Speech	Stem
They	[Pron]	they
ground	[V]	grind
coffee	[Nn]	coffee
every	[Det]	every

Input	Part of Speech	Stem
morning	[Nn]	morning

4.5.3 Guessing Part-of-Speech for Unknown Words

Words not found in the tagger dictionary are passed to the relevant guesser to be assigned the most likely tag. The guesser assigns tags to unknown words based on a set of rules about the morphology of the given language.

Capitalization information may also be used as capitalized words are also proper nouns in many languages. Combinations of alphabetic, numeric and optionally, punctuation characters tend to be guessed as proper nouns as well. Ordinal numbers are tagged either as noun or adjective, depending on the context. Internet and e-mail addresses are assigned the tag `Nn-Net`.

In the Asian languages, unknown words are assigned the tag `Nn` by default.

4.5.4 Language-Specific Unknown Words Guessing Examples

A collection of examples of guessing the part-of-speech for unknown words for certain languages.

4.5.4.1 Arabic Unknown Words

Words not found in the tagger dictionary are passed to the Arabic guesser where they are assigned a tag based on a set of rules about Arabic morphology. The following set of tagging rules are part of this module:

- Feminine nominal and adjectival tags are assigned to words ending with a taa marbuta. A feminine nominal tag is assigned to words ending with `ت`.
- A masculine nominal tag is assigned to words ending with `ين`.
- A masculine adjectival tag is assigned to words ending with `ي`.
- Words that appear to begin with a conjunction, the determiner `ال`, or both, can also be assigned an adjectival or nominal tag.
- An imperfect verb tag is assigned to words beginning with `ت` or `ي`. A perfect verb tag is assigned to words ending with `ت` (but not `ات`).

Unfound words are also assigned the proper noun tag.

4.5.4.2 Catalan Unknown Words

Words not found in the tagger dictionary are passed to the Catalan guesser where they are assigned a tag based on a set of rules about Catalan morphology and capitalization.

The following set of tagging rules are part of the Catalan module.

- Verb tags are assigned according to the verb conjugation patterns. Internet and e-mail addresses are tagged as `Nn-Net`.
- Words beginning with a capital letter or a number followed by a capital letter are guessed as proper nouns. Combinations of alphabetic, numeric and optionally, punctuation characters are also guessed as proper nouns. Combinations of digits and punctuation are tagged as numbers. A series of punctuation marks is tagged as punctuation.

4.5.4.3 Czech Unknown Words

Words not found in the tagger dictionary are passed to the Czech guesser to be assigned the most likely tag.

Czech guesser assigns tags to unfound words based on a set of rules about Czech morphology. For example, a word ending in **-ova** is likely an adjective. Internet and e-mail addresses are assigned the tag `Nn-Net`.

Capitalization information is also important; for instance, capitalized words tend to be guessed as proper nouns.

4.5.4.4 English Unknown Words

Words not found in the tagger dictionary are passed to the English tagger guesser to be assigned the most likely tag. The English tagger guesser assigns tags to unknown words based on a set of rules about English morphology, for example, a word ending in **-ly** is likely an adverb. Internet and e-mail addresses are assigned the tag `Nn-Net`.

Capitalization information is also important; for instance, capitalized words tend to be guessed as proper nouns. Combinations of alphabetic and numeric characters are guessed as proper nouns as well. Ordinal numbers are tagged either as noun or adjective, depending on the context as determined by the software.

4.5.4.5 French Unknown Words

Words not found in the tagger dictionary are passed to the French tagger guesser where they are assigned a tag based on a set of rules about French morphology and capitalization.

The following set of tagging rules are part of this module.

- Verb tags are assigned according to the verb conjugation patterns. The adverb tag is assigned to words ending in **-ement**, **-amment**, **-emment**, **-iment**. Words ending in **-able(s)**, **-ible(s)**, **-eux**, **-ois** are guessed as adjectives, and words ending in **-gé(s)**, **-ré(s)** as past participles.

- Every other lowercase all-alpha word (not ending in an **-s**) is guessed as a singular noun, lowercase all-alpha words ending in **-s**, **-aux**, and **-men** are guessed as plural nouns, and lowercase all-alpha words ending in **-x**, **-z**, **-ais**, **-ois** are guessed as invariant nouns. Internet and e-mail addresses are tagged as Nn-Net.
- Words beginning with a capital letter or a number followed by a capital letter are guessed as proper nouns. The remainder of the word may also contain numbers, lowercase or uppercase letters, hyphen or slash. Combinations of digits and punctuation are tagged as numbers. A series of punctuation marks is tagged as punctuation.

4.5.4.6 German Unknown Words

Words not found in the tagger dictionary are passed to the German tagger guesser where they are assigned a tag based on a set of rules about German morphology and capitalization.

The following set of tagging rules are part of this module.

- Noun tags are assigned to words ending in a number of nominal suffixes. Verb tags are assigned to lowercase words ending in **-ier** and other specified endings. Adverb tags are assigned to words ending in **-weise**, **-ens**, and **-mal**.
- Words with endings like **-ig**, **-isch**, and **-los** are guessed as adjectives. Internet and e-mail addresses are tagged Nn-Net.
- Capitalized words are guessed to be nouns. These words may contain slashes, numbers, or uppercase letters in the middle (**TelCo**, **Tel/Fax** and **3Com**), but not hyphens or apostrophes. Lowercase words are guessed as adjectives or adverbs. Combinations of punctuation are guessed as punctuation. Combinations of numbers and punctuation are guessed as numbers.

4.5.4.7 Greek Unknown Words

Words not found in the tagger dictionary are passed to the Greek tagger guesser to be assigned the most likely tag based on the Greek morphology.

4.5.4.8 Korean Unknown Words

Words not found in the tagger dictionary are passed to the Korean tagger guesser where they are assigned a tag based on a set of rules about the language morphology.

4.5.4.9 Serbian (Cyrillic & Latin) Unknown Words

Words not found in the tagger dictionary are passed to the Serbian tagger guesser to be assigned the most likely tag.

The Serbian guesser assigns tags to unknown words based on a set of rules about Serbian morphology. For example, a word ending in (Cyrillic) `ити` or (Latin) `iti` is likely an infinitive verb. Internet and e-mail addresses are assigned the tag `Nn-Net`. Capitalization information is also important; for instance, capitalized words tend to be guessed as proper nouns.

4.5.4.10 Slovak Unknown Words

Words not found in the tagger dictionary are passed to the Slovak tagger guesser to be assigned the most likely tag.

The Slovak guesser assigns tags to unknown words based on a set of rules about Slovak morphology. For example, a word ending in `ť` is likely an infinitive verb. Internet and e-mail addresses are assigned the tag `Nn-Net`.

Capitalization information is also important; for instance, capitalized words tend to be guessed as proper nouns.

4.5.4.11 Slovenian Unknown Words

Words not found in the tagger dictionary are passed to the Slovenian tagger guesser to be assigned the most likely tag.

The Slovenian guesser assigns tags to unknown words based on a set of rules about Slovenian morphology. For example, a word ending in `-ti` is likely an infinitive verb. Internet and e-mail addresses are assigned the tag `Nn-Net`.

Capitalization information is also important; for instance, capitalized words tend to be guessed as proper nouns.

4.5.4.12 Spanish Unknown Words

Words not found in the tagger dictionary are passed to the Spanish tagger guesser where they are assigned a tag based on a set of rules about Spanish morphology and capitalization.

The following set of tagging rules are part of this module.

- Verb tags are assigned according to the verb conjugation patterns. Internet and e-mail addresses are tagged as `Nn-Net`.
- Words beginning with a capital letter or a number followed by a capital letter are guessed as proper nouns. Combinations of alphabetic, numeric and optionally, punctuation characters are also guessed as proper nouns. Combinations of digits and punctuation are tagged as numbers. A series of punctuation marks is tagged as punctuation.

4.5.5 Language-Specific Part-of-Speech Tagging Examples

A collection of examples of part-of-speech tagging for certain languages.

4.5.5.1 Arabic Part-of-Speech Tagging

Shows the tag set available for defining Arabic custom entities. The tag names are accompanied by a brief description and one or more examples. Note that the extraction process recognizes both Arabic and Latin numbers.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	ان بي سي
Adj	Adj-ConjDet-Fem	Feminine adjective with attached conjunction and determiner	والاقتصادية
	Adj-Det-Fem	Feminine adjective with attached determiner	المتحدة
	Adj-ConjDet-Masc	Masculine adjective with attached conjunction and determiner	والتلفزيوني
	Adj-Det-Masc	Masculine adjective with attached determiner	العقاري
	Adj-Conj-Fem	Feminine adjective with attached conjunction	والارهابية

Umbrella Tag	Complete Tag	Description	Examples
	Adj-Fem	Feminine adjective	اميركية
	Adj-Conj-Masc	Masculine adjective with attached conjunction	ونهايب
	Adj-Masc	Masculine adjective	احياء
Adv	Adv	Adverb	امس
Conj	Conj	Detached conjunction	و / أو
Det	Det	Detached determiner	ال
Fw	Fw	Function word	إن
Interj	Interj	Interjection	يا
Latin	Latin	Any series of Latin-based letters	Hello text ABC
Nn	Nn-ConjDet-Fem	Feminine noun with attached conjunction and determiner	والدبابات
	Nn-Det-Fem	Feminine noun with attached determiner	الرسالة
	Nn-ConjDet-Masc	Masculine noun with attached conjunction and determiner	والمصالح
	Nn-Det-masc	Masculine noun with attached determiner	الأخلاق
	Nn-Conj-Fem	Feminine noun with attached conjunction	فزيارتك
	Nn-Fem	Feminine noun	مخيمات
	Nn-Conj-Masc	Masculine noun with attached conjunction	وتوقيت
	Nn-Masc	Masculine noun	صديقاً
	Nn-Phone	Phone indicator	هاتف
	Num	Num	Number
Num-Date		Numeric date	2005/5/23

Umbrella Tag	Complete Tag	Description	Examples
	Num-Decimal	Decimal number	100.1143.5
	Num-Time	Numeric time expression	10:20:11 23:00
Part	Part-Fut	Future particle	سوف
	Part-Int	Interrogative particle	هل
	Part-Neg	Negation particle	لما
Prep	Prep	Preposition	الى
Pron	Pron-1P	First person pronoun	ونحن
	Pron-2P	Second person pronoun	بك / انت
	Pron-3P-Fem	Third person feminine pronoun	هي / انها
	Pron-3P-Masc	Third person masculine pronoun	هو / فيه
	Pron-3P	Third person pronoun that can be either masculine or feminine	هما / معهما
	Pron-Dem-Fem	Demonstrative feminine pronoun	هذه / وهذي
	Pron-Dem-Masc	Demonstrative masculine pronoun	ذي / هذا
	Pron-Rel	Relative pronoun	الذي / اللتان
Prop	Prop	Proper noun	إنسايث
	Prop-Email	E-mail address	info@netscape.com
	Prop-URL	URL address	www.netscape.com
Punct	Punct	Punctuation	.,:
	Punct-Close	Closing punctuation)]]>
	Punct-Comma	Comma	،
	Punct-Open	Opening punctuation	(([{<
	Punct-Percent	Percent sign	%

Umbrella Tag	Complete Tag	Description	Examples
	Punct-Quote	Quote sign	" "
	Punct-Sent	Sentence ending punctuation	.
	Punct-Symbol	Special punctuation symbols	/ * \$
V	V-Imperativ	Imperative verb	إسع
	V-Imperf	Imperfect verb	تستوجب
	V-Perf	Perfect verb	وقال

The noun and pronoun tags generally do not include number information such as singular, plural or dual. Therefore, the words معلمين, معلم, and معلمان are all tagged as Nn-Masc.

Also, in Arabic, certain prefixes may appear attached to a noun or adjective. The tags for these tokens include information on these prefixes. Therefore, الاتحاد and الاقتصادية will be tagged as Nn-Det-Masc and Adj-Det-Fem respectively.

Finally, verbs are tagged based on the form used, rather than the actual tense they may represent within the context of the sentence. Therefore, if the verb is formed based on the perfective, it will be tagged as V-Perf, and not as Present Perfect, Simple Past, or Future, for example.

4.5.5.2 Catalan Part-of-Speech Tagging

Shows the tag set available for defining Catalan custom entities. The tag names are accompanied by a brief description and one or more examples. The tag set makes no distinction for gender. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	etc, Cia.
Adj	Adj	Adjective, invariant for number	alfin, antidroga
	Adj-Ord-Pl	Plural, spelled out ordinal adjective	segons, primeres
	Adj-Ord-Sg	Singular, spelled out ordinal adjective	vint-i-tresena
	Adj-Pl	Plural adjective	factibles
	Adj-Sg	Singular adjective	absurd, calent, capaç

Umbrella Tag	Complete Tag	Description	Examples
Adv	Adv	Adverb	ací, abans, gairebé, fins
	Adv-Deg	Adverb that can modify an adjective	bastant, força, gaire, massa, més, mig, molt, poc, prou, tan, tot
	Adv-Int	Interrogative adverb	quan, on, a on, d'on, com, per què
	Adv-Rel	Adverbial relativizer	quan, com, on
Aux	Aux-Inf-be	Infinitive ser	ser
	Aux-Inf-have	Infinitive haver	haver
	Aux-anar	Auxiliary anar	vaig
	Aux-be	Auxiliary ser	serà
	Aux-have	Auxiliary haver	ha, han
Conj	Conj	Conjunction	si, perquè, mentre
	Conj-Coord	Coordinating conjunction	i, o, ni
	Conj-com	Conjunction tant el president com el primer secretari	com
	Conj-que	Conjunction que	que
Det	Det-Def-Pl	Plural definite determiner	les, els
	Det-Def-Sg	Singular definite determiner	l'
	Det-Dem-Pl	Plural demonstrative determiner	aqueixes
	Det-Dem-Sg	Singular demonstrative determiner	aquest, això
	Det-Indef-Pl	Plural indefinite determiner or pronoun	gaire, molts, uns
	Det-Indef-Sg	Singular indefinite determiner or pronoun	bastant, gaire, quant, tant, molt, poc
	Det-Int-Pl	Plural interrogative determiner	quins

Umbrella Tag	Complete Tag	Description	Examples
	Det-Int-Sg	Singular interrogative determiner	quin, quant
	Det-Poss-Pl	Plural possessive determiner	nostres, seues, llurs
	Det-Poss-Sg	Singular possessive determiner	teu, ma, llur
	Det-Rel-Sg	Singular relative determiner	qual
	Det-Rel-Pl	Plural relative determiner	quals
Interj	Interj	Interjection	ui!, eh?
Nn	Nn	Noun, invariable for number	atles, albatros, focus
	Nn-Net	URL or e-mail address	www.inxight.com, info@inxight.com
	Nn-Pl	Plural noun	organitzacions, xarxes, casos, drets
	Nn-Sg	Singular noun	manera, exemple
Num	Num	Numeric expression, or cardinal number	2001, milions, dos
	Num-Ord	Ordinal number	1r, 2n, 3r, 4t, 5è
Part	Part-Neg	The negation particle no	no
Prep	Prep	Preposition	amb, a causa de, darrera, en
	Prep-Det-a	Combination a and determiner	al, als
	Prep-Det-de	Combination de and determiner	del, dels, des del
	Prep-Det-per	Combination per and determiner	pel, pels
	Prep-a	Preposition a	a
	Prep-de	Preposition de	de, d'
	Prep-per	Preposition per	per
Pron	Pron	Pronoun	jo, tu, ell, això

Umbrella Tag	Complete Tag	Description	Examples
	Pron-Adv	Adverbial pronoun	en, hi, n', -en
	Pron-Clitic	Clitic pronoun	s', 'ns, -hi
	Pron-Dem	Demonstrative pronoun	aquests
	Pron-Indef	Indefinite pronoun	moltes
	Pron-Int	Interrogative and exclamative pronoun	qui, què, quant, quantes
	Pron-Oblq	Oblique pronoun	en, ho, ell, em
	Pron-Ord	Ordinal pronoun	tercer
	Pron-Poss	Possessive pronoun	el meu, la seva
	Pron-Rel	Relative pronoun	que, qui, què, qual
	Pron-es	es pronoun	es, se, s', -s
Prop	Prop	Proper noun or alpha numeric combination	Europa, FAO/OMS
Punct	Punct	Other punctuation	:"'{}&/
	Punct-Close	Closed parenthesis)
	Punct-Comma	Comma	,
	Punct-Open	Open parenthesis	(
	Punct-Sent	Sentence ending punctuation	.!?
V	V-Fin	Finite verb	reclamen, reconeix, passa, va
	V-Impv	Imperative verb	satisfacin, tracta
	V-Inf	Infinitive verb	arribar, mantenir, buscar
	V-PrPart	Present participle verb	creant, essent, donant
	V/Adj-PaPart-Pl	Plural past participle verb or adjective	elegits, encaminades
	V/Adj-PaPart-Sg	Singular past participle verb or adjective	fet, assenyalat, mancada

4.5.5.3 Chinese (Simplified) Part-of-Speech Tagging

Shows the tag set available for defining both Traditional and Simplified Chinese custom entities. The tag names are accompanied by a brief description and one or more examples. Simplified Chinese examples are given in GB encoding.

Umbrella Tag	Complete Tag	Description	Simplified Chinese Examples (GB)
Adj	Adj	Adjective	一流, 大型
Adv	Adv	Adverb	仅仅, 非常
	Adv-BAN	Metaphor marker	般, 似的
	Adv-Comp	Comparative adverb	最
	Adv-DENG	Post-nominal abbreviation	等
	Adv-Idiom	Idiomatic expression	寸草春晖, 游人止步
Asp	Asp	Postverbal aspect marker	了, 过, 着
Aux	Aux	Auxiliary verb	应当, 能
Cl	Cl	Classifier	张, 副
Conj	Conj	Clausal conjoiner	不论, 即使
	Conj-Nn	Noun conjoiner	及, 和
Det	Det	Determiner	这, 每, 任何
Interj	Interj	Interjection	哇, 喂
Nn	Nn	Common noun	东西, 菜单, 椅子
	Nn-Ascii	ASCII character noun	a, B
	Nn-Loc	Locative noun	上, 以内, 之中
	Nn-Net	URL or e-mail address	www.inxight.com
	Nn-Prop	Proper noun	香港, 叶尔钦
	Nn-Time	Nominal time expression	今天, 周一, 上半年, 下午
Num	Num	Number	万, 3, 5
Ord	Ord	Ordinal prefix	第

Umbrella Tag	Complete Tag	Description	Simplified Chinese Examples (GB)
Part	Part	Sentence-final particle	吧, 吗
Prep	Prep	Preposition	根据, 以, 由
	Prep-Assoc	Modification marker	的
	Prep-Assoc-ZHI	Noun-modification marker	之
	Prep-Assoc-DI	Verb-modification marker	地
	Prep-Assoc-DEI	Modification marker	得
Pron	Pron	Pronoun	她, 我, 你
Punct	Punct	Punctuation	..., -, ;, :
	Punct-Comma	Comma	,
	Punct-Open	Opening punctuation	(, {, 【
	Punct-Close	Closing punctuation), }, 】,]
	Punct-Sent	Sentence-ending punctuation	。
Quant	Quant	Quantifier	整个, 众多
Verb	Verb	Verb	走, 下雨, 负责

4.5.5.4 Chinese (Traditional) Part-of-Speech Tagging

Shows the tag set available for defining both Traditional and Simplified Chinese custom entities. The tag names are accompanied by a brief description and one or more examples. Traditional Chinese examples are given in Big5 encoding.

Umbrella Tag	Complete Tag	Description	Traditional Chinese Examples (Big5)
Adj	Adj	Non-predicative adjective	一流, 大型
Adv	Adv	Adverb	僅僅, 非常
	Adv-BAN	Metaphor marker	般, 似的

Umbrella Tag	Complete Tag	Description	Traditional Chinese Examples (Big5)
	Adv-Comp	Comparative adverb	最
	Adv-DENG	Post-nominal abbreviation	等
	Adv-Idiom	Idiomatic expression	寸草春暉, 游人止步
Asp	Asp	Postverbal aspect marker	了, 過, 著
Aux	Aux	Auxiliary verb	應當, 能
Cl	Cl	Classifier	張, 副
Conj	Conj	Clausal conjoiner	不論, 即使
	Conj-Nn	Noun conjoiner	及, 和
Det	Det	Determiner	這, 每, 任何
Interj	Interj	Interjection	哇, 喂
Nn	Nn	Common noun	東西, 菜單, 椅子
	Nn-Ascii	ASCII character noun	a, B
	Nn-Loc	Locative noun	上, 以內, 之中
	Nn-Net	URL or e-mail address	www.inxight.com
	Nn-Prop	Proper noun	香港, 葉爾欽
	Nn-Time	Nominal time expression	今天, 周一, 上半年, 下午
Num	Num	Number	萬, 3, 5
Ord	Ord	Ordinal number	第
Part	Part	Sentence-final particle	吧, 矣
Prep	Prep	Preposition	根據, 以, 由
	Prep-Assoc	Modification marker	的
	Prep-Assoc-ZHI	Noun-modification marker	之
	Prep-Assoc-DI	Verb-modification marker	地
	Prep-Assoc-DEI	Modification marker	地

Umbrella Tag	Complete Tag	Description	Traditional Chinese Examples (Big5)
Pron	Pron	Pronoun	她,我,你
Punct	Punct	Punctuation	…, -, ;, :
	Punct-Comma	Comma	,
	Punct-Open	Opening punctuation	(, {, 【
	Punct-Close	Closing punctuation), }, 】
	Punct-Sent	Sentence-ending punctuation	。
Quant	Quant	Quantifier	整個,眾多
Verb	Verb	Verb	走,下雨,負責

4.5.5.5 Croatian Part-of-Speech Tagging

Shows the tag set available for defining Croatian custom entities. The tag names are accompanied by a brief description and one or more examples. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	HDZ, RH
Adj	Adj	Absolutive adjective	dobar, velik
	Adj-Comp	Comparative adjective	bolji, veći
	Adj-Sup	Superlative adjective	najbolji, najveći
Adv	Adv	Absolutive adverb	brzo, mnogo
	Adv-Comp	Comparative adverb	brže, više
	Adv-Sup	Superlative adverb	najbrže, najviše
Conj	Conj-Co	Coordinate conjunction	a, i
	Conj-Sub	Subordinate conjunction	jer, da
Enum	Enum	Enumeration	etc.

Umbrella Tag	Complete Tag	Description	Examples
Interj	Interj	Interjection	hej, jao
Nn	Nn-Acc	Accusative noun	žene, profesori
	Nn-Case	Noun case other than nominative and accusative	ženama, profesorom
	Nn-Nom	Nominative noun	žene, profesor
Num	Num	Numeral	dvanaest, sedamdeset
	Num-Acc	Accusative numeral	jednu, jednog
	Num-Card	Cardinal numeral	tri, četiri
	Num-Case	Numeral case other than nominative and accusative	jednom, dvama
	Num-Nom	Nominative numeral	jedan, dva
	Num-Ord	Numeral ordinal	prvi, drugi
	Prep	Prep	Preposition
Pron	Pron	Pronoun	obje, vas, ovi, moji, koja
	Pron-Pers	Personal pronoun	ja, ti
	Pron-Poss	Possessive pronoun	tvoji, naši
	Pron-Ref	Reflexive pronoun	se
Prop	Prop	Name of a person or thing	Zagreb
Punct	Punct	Punctuation	. : ; -
	Punct-Close	Closing punctuation mark)
	Punct-Comma	Comma	,
	Punct-Open	Opening punctuation mark	(
	Punct-Sent	Sentence-ending punctuation mark	. ! ?
V	V-Aux-Clit	Verb auxiliary clitic	je, sam
	V-Fin	Verb, finite	radimo, nose, nosi

Umbrella Tag	Complete Tag	Description	Examples
	V-Inf	Verb, infinitive	raditi, nosi
	V-Part	Verb, participle	mislecí, uzimající, nosili, no-sio

4.5.5.6 Czech Part-of-Speech Tagging

Shows the tag set available for defining Czech custom entities. The tag names are accompanied by a brief description and one or more examples. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj	Adjective	úřední, úředními, úředním
	Adj-Comp	Comparative adjective	úradnějši, úradnějšími, úradnějším
	Adj-Sup	Superlative adjective	nejuřednějši, nejuřednějším, nejuřednějšími
Adv	Adv	Adverb	úředně, zúředněně, zdeúředně
	Adv-Comp	Comparative adverb	úředněji, zúředněněji, zdeúředněji
	Adv-Sup	Superlative adverb	nejúředněji, nejzúředněněji, nejzdeúředněji
Conj	Conj	Conjunction	či, čili, že
Interj	Interj	Interjection	úhuhu, ó, ólala
Nn	Nn	Invariant noun	sec, pH, um.
	Nn-Pl-Gen	Plural, genitive noun	úředníků, úřeků, úřezů
	Nn-Pl-Case	Plural, nominative, vocative, accusative, dative, locative and instrumental noun	úředníci, úředníkům, úředníků
	Nn-Sg-Gen	Singular, genitive noun	úředníka, úřeku, úřezu

Umbrella Tag	Complete Tag	Description	Examples
	Nn-Sg-Case	Singular, nominative, vocative, accusative, dative, locative and instrumental noun	úředníka, úředníkovi, úředníkem
	Nn-Net	URL, e-mail address	www.inxight.com, info@inxight.com,
Prop	Prop	Proper noun	ČSLA, Ňasko, Ňasku, Ňaska, Ňaskem, Íliada
Num	Num	Number expression other than cardinal or ordinal	XV, mil.
	Num-Card	Cardinal number	dvě, dvěma, dvou
	Num-Ord	Ordinal number	šestýma, šestými, šestým
Part	Part	Particle	řekněmež, čau, žbluňk
Prep	Prep	Preposition	zmísta, zkraje, zaň
Pron	Pron-Dem-Pl	Plural demonstrative pronoun	týmiž, týmaž, týchž
	Pron-Dem-Sg	Singular demonstrative pronoun	týž, týmž
	Pron-Pl	Plural pronoun	číchkoliv, číchkoli, čímisi, čímasi, čímsi
	Pron-Sg	Singular pronoun	číhosi, čímisi, čímukoliv, čímukoli, číhokoliv, číhokoli, číhosi
	Pron-Int/Rel	Interrogative/relative pronoun	čí, čími, čím, čích, čímú
	Pron-Refl	Reflexive pronoun	svůj, svých, svému, svým
	Pron-Pers-Sg	Singular personal pronoun	on, ono, ona, ty, von
	Pron-Pers-Pl	Plural personal pronoun	vy, vás, vám, vámi
	Pron-Poss	Possesive pronoun	tvůj, váš, vaší
Punct	Punct-Sent	Sentence ending punctuation	! ? .
	Punct-Comma	Comma	,

Umbrella Tag	Complete Tag	Description	Examples
	Punct-Open	Opening punctuation	(
	Punct-Close	Closing punctuation)
	Punct-Quote	Quote	" "
	Punct	Other punctuation	+ -
V	V-Inf	Infinitive verb	dělat, užít, užívat
	V-Imp	Imperative verb	dělej, dělejme, dělejte
	V-Ind	Indicative, verb	dělána, dělány, dělání, dělání, dělání, dělání
	V-PaPart	Past participle	dělal, dělals, dělaly, dělali
	V-Inf-Be	Verb "to be", infinitive	být, bývat, nebýt, nebývat
	V-Imp-Be	Verb "to be", imperative mood	buď, budiž, budme, budte
	V-Pres-Be	Verb "to be", present tense	je, jest, jsi, jste, jsme, jsou, jsem
	V-Fut-Be	Verb "to be", future tense	bude, budu, budeš, budete, budou, budem
	V-PaPart-Be	Verb "to be", past participle	byl, byla, bylo, byla, byly, byli
	V-APart	Adjectival/adverbial participle	dělaje, dělajíc, dělající
	V-Aux	Auxiliary verb	by, bys, byste, bych, bychom

4.5.5.7 Danish Part-of-Speech Tagging

Shows the tag set available for defining Danish custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	nov., kg, dkr., USA

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj	Adjective	det gode brød
	Adj-Comp	Comparative adjective	et bedre forslag
	Adj-Sup	Superlative adjective	det bedste forslag
	Adj-Gen	Genitive adjective	den enkeltes tryghed
	Adj-PaPart	Past participle used as adjective	bestemt niveau
	Adj-PaPart-Gen	Past participle used as adjective, genitive	den ansattes papirer
	Adj-PrPart	Present participle adjective	manglende
	Adj-Sup	Superlative adjective	det bedste forslag
Adv	Adv	Adverb (includes particles)	igen
	Adv-Comp	Comparative adverb	tidligere
	Adv-Int/Rel	Interrogative or relative adverb	hvor, hvorefter
	Adv-Sup	Superlative adverb	oftest
Aux/V	Aux/V-Infin-blive	Infinitive auxiliary or main verb blive	blive
	Aux/V-Infin-faa	Infinitive auxiliary or main verb få	få
	Aux/V-Infin-have	Infinitive auxiliary or main verb have	have
	Aux/V-Infin-vaere	Infinitive auxiliary or main verb vaere	være
	Aux/V-PaPart-blive	Past participle auxiliary or main verb blive	blevet
	Aux/V-PaPart-faa	Past participle auxiliary or main verb få	fået
	Aux/V-PaPart-have	Past participle auxiliary or main verb have	haft

Umbrella Tag	Complete Tag	Description	Examples
	Aux/V-PaPart-vaere	Past participle auxiliary or main verb vaere	været
	Aux/V-Past-blive	Past tense auxiliary or main verb blive	blev
	Aux/V-Past-faa	Past tense auxiliary or main verb få	fik
	Aux/V-Past-have	Past tense auxiliary or main verb have	havde
	Aux/V-Past-vaere	Past tense auxiliary or main verb vaere	var
	Aux/V-Pres-blive	Present tense auxiliary or main verb blive	bliver
	Aux/V-Pres-faa	Present tense auxiliary or main verb få	får
	Aux/V-Pres-have	Present tense auxiliary or main verb have	har
	Aux/V-Pres-vaere	Present tense auxiliary or main verb vaere	er
Cmpd	Cmpd-Part	Left compound part	post - og telegrafvæsenet
Conj	Conj	Conjunction	at, når
	Conj-Coord	Coordinating conjunction	og, eller
	Conj-hvis	Conjunction or relative pronoun hvis	hvis
	Conj-som	Conjunction or relative pronoun som	som
Det	Det	Determiner	en
	Det-Indet	Indeterminate determiner	forskellig, somme
	Det-Indet-Gen	Indeterminate determiner, genitive	forskelliges
	Det-Coord	Conjunctive adverb	både

Umbrella Tag	Complete Tag	Description	Examples
	Det/Pron-Int/Rel	Interrogative or relative pronoun	hvad, hvem, hvilke
	Det/Pron-Poss	Possessive determiner or pronoun	vores, min
	Det/Pron-Poss-Refl	Reflexive possessive pronoun	sin, sit, sine
	Det/Pron-Quant	Quantifying determiner or pronoun	mange
	Det/Pron-Quant-Comp-mere	Comparative mere	mere
	Det/Pron-Quant-Gen	Genitive quantifying determiner or pronoun	manges
	Det/Pron-Quant-Pre	Quantifying pre-determiner or pronoun	alle, hver
	Det/Pron-Quant-Sup-mest	Superlative mest	mest
Interj	Interj	Interjection	hej
Nn	Nn	Noun	kvinde
	Nn-Gen	Genitive noun	kvindens
	Nn-Letter	Lowercase and uppercase letters	b, N
	Nn-Net	URL and e-mail address	www.inxight.com info@inxight.com
Num	Num	Cardinal number (in digits or words)	3m tre
Ord	Ord	Ordinal number, in digits or spelled out	20., femte
Part	Part-Inf	Infinitival particle at	få lov at indtage
	Part-Neg	Negative particle	ikke
Prep	Prep	Preposition	med, hos
	Prep-af	Preposition af	af

Umbrella Tag	Complete Tag	Description	Examples
Pron	Pron	Pronoun	den, denne
	Pron-Expl	Expletive pronoun	der var 400 deltagere
	Pron-Gen	Genitive pronoun	begges
	Pron-Pers	Personal pronoun	jeg, mig
	Pron-Recip	Reciprocal pronoun	hinanden
	Pron-Recip-Gen	Genitive reciprocal pronoun	hinandens
	Pron-Rel	Relative pronouns der and som	familier, der skilles
Prop	Prop	Proper name, initials or title	Ole, H., fru, dr.
	Prop-Gen	Genitive proper name	Jensens bil
Punct	Punct	Miscellaneous punctuation	-)
	Punct-Comma	Comma	,
	Punct-Sent	Sentence boundary punctuation	. ? !
V	V-Impv	Imperative verb	skriv
	V-Infin	Infinitive verb	skrive
	V-PaPart	Past participle verb	skrevet
	V-Past	Past tense verb	skrev
	V-Past-SForm	Past tense S-form verb	taltes
	V-Pres	Present tense verb	sker
	V-Pres-SForm	Present tense S-form verb	sendes

4.5.5.8 Dutch Part-of-Speech Tagging

Shows tag set available for defining Dutch custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj-Attr	Attributive adjective	een snelle auto
	Adj-Post	Postmodifying adjective	wat anders
	Adj-Pred	Predicative adjective	hij rijdt snel
Adv	Adv	Non-adjectival adverb	stroomopwaarts
	Adv-Deg	Adverb that can modify an adjective	hij wil te snel
	Adv-Int	Interrogative adverb	waarom gaat hij
	Adv-Pron	Pronominal adverb	hij praat hierover
Aux	Aux-Fin	Finite auxiliary verb	hij is geweest
	Aux-Inf	Infinitive auxiliary verb	hij zal zijn
	Aux-PaPart	Past participle auxiliary verb	hij is geweest
Cmpd	Cmpd-Left	Left truncated part of compound	honden- en kattenvoer
	Cmpd-Right	Right truncated part of compound	kattenvoer en -melk
Conj	Conj-Comp	Comparative conjunction	zo groot als
	Conj-Coord	Coordinating conjunction	jan en marie
	Conj-Inf	Infinitive conjunction	alvorens te vragen
	Conj-Rel	Relative conjunction	het kind dat ...
	Conj-Sub	Subordinating conjunction	hoewel hij er was
	Conj-Sub-Adv	Interrogative adverb or subordinate conjunction	wanneer gaat hij weg?
Det	Det-Art	Determiner	een bus
	Det-Dem	Demonstrative determiner	deze machine gaat goed
	Det-Indef	Indefinite determiner	geen broer
	Det-Int/Rel	Interrogative or relative determiner	de vraag wier man ...
	Det-Poss	Possessive determiner	mijn boek

Umbrella Tag	Complete Tag	Description	Examples
	Det-Post-Indef	Indefinite postdeterminer	de beide broers
	Det-Pre-Indef	Indefinite predeterminer	binnen al deze pakketten
Interj	Interj	Interjection	och
Nn	Nn	Common noun	boek
	Nn-Letter	Lowercase and uppercase letters	b, N
	Nn-Net	URL and e-mail address	www.inxight.com info@inxight.com
Num	Num	Cardinal number	125, vijf, 12/2
Ord	Ord	Ordinal number	vijfde, 125ste, 12de
Part	Part-Inf	Particle of Dutch 'te+infinitive' construction	hij hoopt te gaan
	Part-Neg	Negation particle	hij gaat niet snel
	Part-Prefix	Separated prefix of (pronominal) adverb or verb	hij loopt mee
Prep	Prep	Preposition	in
	Prep-Circ	Right part of circumposition	tot nu toe
	Prep-Post	Postposition	veel passanten langs komen
	Prep-van	Preposition van	van
Pron	Pron-Dem	Demonstrative pronoun	deze gaat goed
	Pron-Indef	Indefinite pronoun	beide
	Pron-Int/Rel	Interrogative or relative pronoun	de vraag wie ...
	Pron-Pers	Personal pronoun	hij
	Pron-Rel	Relative pronoun	de man die lachte
Prop	Prop	Proper noun, including initials and title of address	Peter, C., Prof.
	Prop-Art	Article beginning a name	De Vries

Umbrella Tag	Complete Tag	Description	Examples
	Prop-Prep	Preposition beginning a name	Van den Broek
Punct	Punct	Miscellaneous punctuation	{ } [] - ---
	Punct-Comma	Comma	,
	Punct-Quote	Quotation type punctuation (includes parentheses)	'' '()
	Punct-Sent	Sentence final punctuation	. ? ! ;
	Punct-Slash	Slash mark	/
V	V-Fin	Finite verb	zegt
	V-Inf	Infinitive verb	zeggen
	V-PaPart	Past participle verb	gezegd
	V-PrPart	Present participle verb	zeggend

Related Information

[Part-of-Speech Tagging \[page 87\]](#)

4.5.5.9 English Part-of-Speech Tagging

Shows the tag set available for defining English custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	i.e.
Adj	Adj	Adjective	big
	Adj-Comp	Comparative adjective	bigger
	Adj-Ord	Ordinal adjective	third

Umbrella Tag	Complete Tag	Description	Examples
	Adj-Sup	Superlative adjective	biggest
Adv	Adv	Adverb	quickly
	Adv-Comp	Comparative adverb	sooner
	Adv-Int/Rel	wh- adverb	how
	Adv-Sup	Superlative adverb	soonest
Aux	Aux	Auxiliary or modal	could
Conj	Conj-Coord	Coordinating conjunction	and
	Conj-Sub	Subordinating conjunction	unless
Det	Det	Invariant determiner (singular or plural)	some food
	Det-Def	Definite determiner	the
	Det-Indef	Indefinite determiner	an
	Det-Int	Interrogative determiner	what time?
	Det-Int/Rel	Interrogative or relative determiner	whose
	Det-Pl	Plural determiner	those apples
	Det-Poss	Possessive determiner	my
	Det-Rel	Relative determiner	whatsoever
	Det-Sg	Singular determiner	every
Interj	Interj	Interjection	oh, hello
Nn	Nn	Invariant noun	sheep
	Nn-Letter	Letter	b, N
	Nn-Net	URL, e-mail address	www.inxight.com, info@inxight.com
	Nn-Pl	Plural noun	computers
	Nn-Sg	Singular noun	farmer

Umbrella Tag	Complete Tag	Description	Examples
Num	Num	Cardinal number or other numeric expression	40.5, 11/27/00, \$12.55, 12%, xvii, 9:00
Part	Part-Inf	Infinitive marker	to be or not to be
	Part-Neg	Negative particle	not
	Part-Poss	Possessive marker	John's coat
Prep	Prep	Preposition	below
	Prep-at	Preposition at	at
	Prep-of	Preposition of	of
Pron	Pron	Pronoun	he
	Pron-Int	wh pronoun	what do you want?
	Pron-Int/Rel	wh pronoun	who
	Pron-Refl	Reflexive pronoun	himself
	Pron-Rel	Relative pronoun	whoever
Prop	Prop	Name of a person or thing	Graceland
Punct	Punct	Miscellaneous punctuation	- ; /%\$
	Punct-Close	Closing punctuation)] }
	Punct-Comma	Comma	,
	Punct-Open	Opening punctuation	([{
	Punct-Quote	Quote	" "
	Punct-Sent	Sentence-ending punctuation	. ! ?
V	V-Inf-be	Infinitive to be	be
	V-PaPart	Verb, past participle, -ed verb form	has walked
	V-PaPart-be	Past participle of to be	has been
	V-PaPart-have	Past participle of to have	he has had

Umbrella Tag	Complete Tag	Description	Examples
	V-Past	Verb, past tense	ran
	V-Past-have	Past tense of have	we had
	V-Past-Pl-be	Verb, past tense plural of to be	were
	V-Past-Sg-be	Verb, past tense singular of to be	was
	V-Pres	Verb, present tense or infinitive	sit
	V-Pres-3-Sg	Verb, present tense, 3rd person singular	sits
	V-Pres-3-Sg-have	Present tense, 3rd person singular of have	has
	V-Pres-have	Present tense or infinitive of have	have
	V-Pres-Pl-be	Verb, present tense plural of to be	are
	V-Pres-Sg-be	Verb, present tense singular of to be	is
	V-PrPart	Verb, present participle, -ing verb form	is walking

4.5.5.10 Farsi Part-of-Speech Tagging

Shows the tag set available for defining Farsi custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	ه.ق.
Acc	Acc	Accusative marking on direct objects	را

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj-Comp	Comparative singular adjective	بزرگتر
	Adj-NPB	Singular adjective, marks NP boundary	زیبا
	Adj-NoNPB	Singular adjective, not an NP boundary	زیبای
	Adj-Pl	Plural adjective	بزرگان
	Adj-Pl-Comp	Comparative plural adjective	بزرگتران
	Adj-Pl-NPB	Plural adjective, marks NP boundary	زیباها
	Adj-Pl-NoNPB	Plural adjective, not an NP boundary	زیباهای
	Adj-Pl-Sup	Superlative plural adjective	کوچکترینان
	Adj-Sup	Superlative singular adjective	بزرگترین
Adv	Adv	Adverb	همیشه
Conj	Conj	Conjunction	تا آنجائیکه
Coord	Coord	Coordination	و
Det	Det	Determiner	این
Nn	Nn	Singular noun	کتاب
	Nn-Day	Day noun	دوشنبه
	Nn-Mon	Month noun	فروردین
	Nn-NPB	Noun, marks NP boundary	کتابی
	Nn-NonNPB	Noun, not an NP boundary	دانشجوی
	Nn-Pl	Plural noun	کارخانجات
	Nn-Pl-NPB	Plural noun, marks NP boundary	کتاب ها
	Nn-Pl-NoNPB	Plural noun, not an NP boundary	کتاب های

Umbrella Tag	Complete Tag	Description	Examples
Num	Num	Number	چهار
	Num-Ord	Ordinal number	چهارمین
Phrs	Phrs	Phrase	به به
Prep	Prep	Preposition	در
Pron	Pron	Pronoun	ما
Prop	Prop	Proper noun	داریوش
Punct	Punct	Punctuation	,
	Punct-Sent	Sentence boundary punctuation	؟
Qua	Qua	Quantifier	همه
Que	Que	Question word or wh-phrase	کی
Rel	Rel	Relativizer	که
Title	Title	Title	آقا
V	V-Fut	Future verb	خواهم رفت
	V-Impv	Imperative verb	برویم
	V-Inf	Infinitive verb	رفتن
	V-Part	Past participle or adjectival verb	رفته
	V-Past	Past verb	رفته بودیم
	V-PrPast	Present participle verb	خندان
	V-Pres	Present verb	می رویم

Boundary markers

The `NPB` and `NONPB` tags signify NounPhraseBoundary and NoNounPhraseBoundary, respectively. These tags are used to indicate whether a boundary defining affix is present on the noun or adjective. Since phrasal boundaries are highly ambiguous in Farsi, this information may be useful for parsing applications that need to distinguish noun phrases and preposition phrases.

Hence, nouns and adjectives ending in the pronoun affix (for example, همسایگانشان 'their neighbors'), the copula verb (مسلمانند 'they are muslims'), the indefinite marker (کتابهایی '(some) books') or the object marker (رئیس جمهور را 'the President'), as well as those ending in a vowel (زیبا 'beautiful'), would be tagged with the NPB.

On the other hand, nouns and adjectives ending in an **ezafe** affix (for example, دانشجوی 'the university student (of)') would be marked with the N_{ONPB} tag since the **ezafe** indicates that the noun should be linked to the following element in the noun phrase.

Miscellaneous tags

Nouns that are formed from a verbal root such as شنونده (listener) are tagged as Nouns in the Farsi tagger. Adjectival or participial forms such as شنیده (heard) are tagged as V-Part as shown in the tagset above.

Organization names such as اوپک (OPEC) are tagged as Nouns. Similarly, any foreign word written in ASCII is marked with the Nn tag. However, e-mail and Internet addresses are tagged as Prop (proper names).

Cardinal and ordinal numbers are distinguished in the Farsi tagger since they differ in their syntactic behavior and word order relative to the rest of the noun phrase.

4.5.5.11 French Part-of-Speech Tagging

Shows the tag set available for defining French custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj	Invariant adjective	heureux
	Adj-Ord-Pl	Spelled out plural ordinal number	deuxièmes
	Adj-Ord-Sg	Spelled out singular ordinal number	deuxième
	Adj-Pl	Plural adjective	gentilles
	Adj-Pre	Preposed invariant adjective	vieux
	Adj-Pre-Pl	Preposed plural adjective	petits chiens
	Adj-Pre-Sg	Preposed singular adjective	petit chien
	Adj-Sg	Singular adjective	gentille
Adv	Adv	Adverb	finale ^{ment} , aujourd'hui

Umbrella Tag	Complete Tag	Description	Examples
	Adv-Deg	Adverb that can modify an adjective	très
Aux	Aux-Fin-12	1st or 2nd person auxiliary, any tense	fusse
	Aux-Fin-3-Pl	3rd person plural auxiliary, any tense	seraient
	Aux-Fin-3-Sg	3rd person singular auxiliary, any tense	serait
	Aux-Inf	Infinitive auxiliary	en avoir assez
	Aux-PaPart	Past participle auxiliary	eu
	Aux-PrPart	Present participle auxiliary	ayant
Conj	Conj-Adv	Connecting or subordinating conjunction	quand
	Conj-comme	Comme	comme
	Conj-Coord	Coordinating conjunction	et, ou
	Conj-que	Que	que
Det	Det-Pl	Plural determiner	vos
	Det-Sg	Singular determiner	ma, votre
Interj	Interj	Interjection	hi, pouah
Nn	Nn	Invariant noun	taux
	Nn-Letter	Letter	z, K
	Nn-Net	URL and e-mail address	www.inxight.com, info@inxight.com
	Nn-Pl	Plural noun	chiens, fourmis
	Nn-Sg	Singular noun	chien, fourmi
Num	Num	Numeral or digit expression	treize, un million, 12, 15kHz, XIX
Part	Part-Neg	Negation particle	ne

Umbrella Tag	Complete Tag	Description	Examples
	Part-voicila	Particles voici and voilà	voici, voilà
Prep	Prep	Preposition (other than à au de du ...)	dans
	Prep-a	Preposition à	à, au, aux
	Prep-de	Preposition de	de, des, du, d'
	Prep-en	Preposition en	en bonne santé
Pron	Pron	Pronoun	il, elles je
	Pron-12	1st or 2nd person pronoun	je
	Pron-Clit	Clitic pronoun	donne- le , donne- lui
	Pron-IntRel	Relative or interrogative pronoun (except que)	qui, quoi, lequel
Prop	Prop	Proper noun	Marie, Paris
Punct	Punct	Punctuation	:-
	Punct-Comma	Comma	,
	Punct-Quote	Quotation marks	"
	Punct-Sent	Sentence-ending punctuation	. ! ? ;
V/Adj	V/Adj-PaPart	Invariant past participial verb or adjective	souri
	V/Adj-PaPart-Pl	Plural past participial verb or adjective	lues
	V/Adj-PaPart-Sg	Singular past participial verb or adjective	dansé
V	V-Fin-12	1st or 2nd person verb, any tense	dansiez, dansais
	V-Fin-3-Pl	3rd person plural verb, any tense	danseront
	V-Fin-3-Sg	3rd person singular verb, any tense	dansait

Umbrella Tag	Complete Tag	Description	Examples
	V-Inf	Infinitive verb	dans er , fin ir
	V-PrPart	Present participle verb	not ant

4.5.5.12 German Part-of-Speech Tagging

Shows the tag set available for defining German custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj-Attr	Attributive adjective	schwarze Katze
	Adj-Attr-Comp	Comparative attributive adjective	kleinere Datei
	Adj-Attr-Ord	Spelled out ordinal number	drit t er Mann
	Adj-Attr-Sup	Superlative attributive adjective	schnellster Läufer
	Adj-Dem	Demonstrative adjective	sol che
	Adj-Indet	Indefinite adjective	dein ig e
	Adj-Pred	Predicate adjective (may also be an adverb)	schn ell
	Adj-Pred-Comp	Comparative predicate adjective	bess er
	Adj-Pred-Sup	Superlative predicate adjective	am schnellsten
Adv	Adv	Adverb	immer, zw ar
	Adv-Int/Rel	Interrogative or relative adverb	wies o
	Adv-Pron	Pronominal adverb	hierf ür
Aux	Aux-Fin	Finite auxiliary verb	bist
	Aux-Inf	Auxiliary verb infinitive	gebildet haben

Umbrella Tag	Complete Tag	Description	Examples
	Aux-PaPart	Auxiliary verb past participle	gehabt
Cmpd	Cmpd-Left	Left compound part	zwei-, Kontakt-
Conj/Adv	Conj/Adv	Conjunction or adverb	jedoch
Conj	Conj-als	Conjunction als	als
	Conj-Coord	Coordinating conjunction	und
	Conj-Inf	Infinitival conjunction	um ... zu
	Conj-Post	Correlative conjunction	weder ... noch
	Conj-Pre	Preposed conjunction	weder
	Conj-Subord	Subordinating conjunction	weil
	Conj-wie	Conjunction wie	wie
Det/Pron	Det/Pron-Quant	Quantifying determiner or pronoun	lauter Sachen
Det	Det-Art	Article	die, das
	Det-Dem	Demonstrative determiner	diese
	Det-Indet	Indefinite determiner	keiner
	Det-Int/Rel	Interrogative determiner or pronoun	wieviel
	Det-Poss	Possessive determiner	dein
Interj	Interj	Interjection	ach, oh
Modal	Modal-Fin	Finite modal verb	darf
	Modal-Inf	Modal verb infinitive	gehörchen sollen
	Modal-PaPart	Modal verb past participle	segeln gedurft
Nn	Nn	Noun	Tisch
	Nn-Letter	Lowercase letters with or without a period and uppercase letters	Ein e Ein r .

Umbrella Tag	Complete Tag	Description	Examples
	Nn-Net	URL and e-mail address	www.inxight.com, info@inxight.com
Num	Num	Cardinal number or date	zwei, 2.3.1999
	Num-Ord	Ordinal number	43.
Part	Part-Ant	Sentential particle	danke
	Part-Comp	Comparative particle	viel besser
	Part-Inf	Infinitival particle	zu sagen
	Part-Neg	Negation particle	nicht
	Part-Num	Numerical particle	rund 50 Dateien
	Part-Pos	Positive particle	zu schnell
	Part-Pref	Separable prefix	'
	Part-Sup	Superlative particle	am besten
Prep	Prep-aus	Preposition aus	aus
	Prep-Circ	Last part of a circumposition	um Himmels willen
	Prep-Det	Preposition-article combination	zum
	Prep-fuer	Preposition für	für
	Prep-Post	Postposition	dem Haus gegenüber
	Prep-Pre	Preposition	mit
	Prep-pro	Preposition pro	pro
	Prep-von	Preposition von	von
Pron	Pron-Dem	Demonstrative pronoun	diese ist besser
	Pron-Dem-Inv	Uninflected demonstrative	solch ein Erfolg
	Pron-Indet	Indefinite pronoun	niemand
	Pron-Indet-Inv	Uninflected determiner	manch ein Mensch

Umbrella Tag	Complete Tag	Description	Examples
	Pron-Int/Rel	Interrogative or relative pronoun	was, wer
	Pron-Int/Rel-Inv	Uninflected interrogative or relative pronoun	was für
	Pron-Pers	Personal pronoun	ich
	Pron-Poss	Possessive pronoun	meine sind gelb
	Pron-Recip	Reciprocal pronoun	einander
	Pron-Refl	Reflexive pronoun	sich
	Pron-Rel	Relative pronoun	die Leute, die . . .
Punct	Punct	Punctuation	()
	Punct-Comma	Comma	,
	Punct-Sent	Sentence-ending punctuation	. ? !
V	V-Fin	Finite verb	schwimmt
	V-Inf	Infinitive verb	er kann schwimmen
	V-Izu	zu infinitive	auszubilden
	V-PaPart	Past participle verb	er ist geschwommen

4.5.5.13 Greek Part-of-Speech Tagging

Shows the tag set available for defining Greek custom entities. The tag names are accompanied by a brief description and one or more examples.

Umbrella tag	Tag	Description	Examples
Adj	Adj-Inv	Adjective invariant in case and number	καφέ, μίνι, ρετρό, εξπέρ, μπλε
	Adj-Pl-Acc	Plural adjective in accusative case	μέσα, όλα, πολλές, χρόνια, σύμφωνα

Umbrella tag	Tag	Description	Examples
	Adj-Pl-Gen	Plural adjective in genitive case	όλων, κοινωνικών, μεγάλων, υλικών, μέσων
	Adj-Pl	Plural adjective, invariant in case	δύο, δυο, πέντε, δώδεκα, οκτώ
	Adj-Pl-Nom	Plural adjective in nominative case	μέσα, όλα, πολλές, σύμφωνα, χρόνια
	Adj-Pl-Voc	Plural adjective in vocative case	μέσα, όλα, πολλές, σύμφωνα, χρόνια
	Adj-Sg-Acc	Singular adjective in accusative case	μια, ένα, πολύ, μόνο, πρώτη
	Adj-Sg-Gen	Singular adjective in genitive case	ενός, μιας, ελληνικής, προκειμένου, ελληνικού
	Adj-Sg	Singular adjective, invariant in case	ενάμισι, ενάμισι, εναμισι, εναμισι
	Adj-Sg-Nom	Singular adjective in nominative case	μια, ένα, πολύ, μόνο, πρώτη
	Adj-Sg-Voc	Singular adjective in vocative case	πολύ, μόνο, πρώτη, ίδιο, όλο
Adv	Adv	Adverb	που, πως, ως, όπως
Art	Art-Def-Pl-Acc	Plural definite article in accusative case	τα, τους, τις, στα, στις
	Art-Def-Pl-Gen	Plural definite article in genitive case	των, ζτων, στων
	Art-Def-Pl-Nom	Plural definite article in nominative case	τα, οι, τ'
	Art-Def-Sg-Acc	Singular definite article in accusative case	το, την, στο, στην, τον
	Art-Def-Sg-Gen	Singular definite article in genitive case	της, του, τ', στ', στου
	Art-Def-Sg-Nom	Singular definite article in nominative case	το, η, ο, τ'
	Art-Indef-Sg-Acc	Singular indefinite article in accusative case	μια, ένα, μία, έναν, μιν

Umbrella tag	Tag	Description	Examples
	Art-Indef-Sg-Gen	Singular indefinite article in genitive case	ενός, μιας, μίας, ενος, ενός
	Art-Indef-Sg-Nom	Singular indefinite article in nominative case	μια, ένα, μία, ένας, ένας
Conj	Conj-Coord	Coordinating conjunction	και, η, κι, ή, αλλά
	Conj-Sub	Subordinating conjunction	να, που, ότι, πως, όπως
Interj	Interj	Interjection	αχ, ρτοπ, ρουτ, στοπ, σουτ
Nn	Nn-Inv	Noun invariant in number and case	ματς, γκολ, φεστιβάλ, σκορ, σεφ
	Nn-Pl-Acc	Plural noun in accusative case	μέσα, παιδιά, στοιχεία, χρόνια, σύμφωνα
	Nn-Pl-Gen	Plural noun in genitive case	ομάδων, προσώπων, χρόνων, παιδιών
	Nn-Pl	Plural noun, invariant in case	κόμικς, φασαμέν, φασαμεν, φασαμέν, τσιπς
	Nn-Pl-Nom	Plural noun in nominative case	μέσα, παιδιά, στοιχεία, χρόνια, σύμφωνα
	Nn-Pl-Voc	Plural noun in vocative case	μέσα, παιδιά, στοιχεία, χρόνια, σύμφωνα
	Nn-Sg-Acc	Singular noun in accusative case	Ελλάδα, φορά, ζωή, παιδιά, όλο
	Nn-Sg-Gen	Singular noun in genitive case	Ελλάδας, κυβέρνησης, παρά, αιώνα, ζωής
	Nn-Sg	Singular noun invariant in case	είναι, εγώ, πρωί, χιούμορ, αύριο
	Nn-Sg-Nom	Singular noun in nominative case	Ελλάδα, φορά, ζωή, παιδιά, όλο
Nn-Sg-Voc	Singular noun in vocative case	Ελλάδα, φορά, ζωή, παιδιά, όλο	
Num	Num	Numerals	23, 33.1, 4, 22-23
Part	Part	Particle	που, για, μα, ν', ας

Umbrella tag	Tag	Description	Examples
	Part-Fut	Future particle	θα, θ', θά
	Part-Neg	Negative particle	δεν, δε, όχι, μη, μην
Prep	Prep	Preposition	με, από, για, σε, ως
Pron	Pron-Inv	Pronoun invariant in number and case	κάτι, τι, τίποτα, τίποτε, κατι
	Pron-Acc	Pronoun in accusative case invariant in number	κάνα, κανα
	Pron-Nom	Pronoun in nominative case invariant in number	κάνα, κανα
	Pron-Pl-Acc	Pronoun in accusative case, plural	αυτά, αυτές, άλλα, κάποια, άλλες
	Pron-Pl-Gen	Pronoun in genitive case, plural	τους, μας, σας, αυτών, άλλων
	Pron-Pl-Nom	Pronoun in nominative case, plural	αυτά, αυτές, άλλα, κάποια, άλλες
	Pron-Sg-Acc	Pronoun in accusative case, singular	μια, ένα, αυτό, αυτή, τόσο
	Pron-Sg-Gen	Pronoun in genitive case, singular	της, του, μου, σου, ενός
	Pron-Sg	Pronoun invariant in case, singular	κάθε, καθε
	Pron-Sg-Nom	Pronoun in nominative case, singular	μια, ένα, αυτό, αυτή, τόσο
	Pron-Pers-Pl1-Acc	Personal pronoun in accusative case, 1st person plural	μας, μάς, εμάς, εμας, εμάσ
	Pron-Pers-Pl1-Gen	Personal pronoun in genitive case, 1st person plural	μας, μάς, εμάς, εμας, εμάσ
Pron-Pers-Pl1-Nom	Personal pronoun in nominative case, 1st person plural	εμείς, μεις, μεις, εμεις, εμεις	
Pron-Pers-Pl2-Acc	Personal pronoun in accusative case, 2nd person plural	σας, εσάς, ζάς, ζάς, σοσ	

Umbrella tag	Tag	Description	Examples
	Pron-Pers-Pl2-Gen	Personal pronoun in genitive case, 2nd person plural	σας, εσάς, ζάς, ζάς, σας
	Pron-Pers-Pl2-Nom	Personal pronoun in nominative case, 2nd person plural	εσείς, σεις, σεις, εξείς, εξείς
	Pron-Pers-Pl2-Voc	Personal pronoun in vocative case, 2nd person plural	εσείς, σεις, σεις, εξείς, εξείς
	Pron-Pers-Pl3-Acc	Personal pronoun in accusative case, 3rd person plural	τα, τους, τις, αυτά, αυτές
	Pron-Pers-Pl3-Gen	Personal pronoun in genitive case, 3rd person plural	τους, αυτών, αυτων, τους, τούς
	Pron-Pers-Pl3-Nom	Personal pronoun in nominative case, 3rd person plural	τα, αυτά, αυτές, αυτοί, αυτοε
	Pron-Pers-Sg1-Acc	Personal pronoun in accusative case, 1st person singular	με, μ', μένα, εμένα, μέ
	Pron-Pers-Sg1-Gen	Personal pronoun in genitive case, 1st person singular	μου, εμένα, μού, εμενα
	Pron-Pers-Sg1-Nom	Personal pronoun in nominative case, 1st person singular	εγώ, γω, εγω
	Pron-Pers-Sg2-Acc	Personal pronoun in accusative case, 2nd person singular	σε, σ', σένα, εσένα, ζενα
	Pron-Pers-Sg2-Gen	Personal pronoun in genitive case, 2nd person singular	σου, εσένα, σου, εξένα, εξενα
	Pron-Pers-Sg2-Nom	Personal pronoun in nominative case, 2nd person singular	εσύ, συ, ζυ, εξυ, εξύ
	Pron-Pers-Sg2-Voc	Personal pronoun in vocative case, 2nd person singular	εσύ, συ, ζυ, εξυ, εξύ
	Pron-Pers-Sg3-Acc	Personal pronoun in accusative case, 3rd person singular	το, την, τον, τη, αυτό
	Pron-Pers-Sg3-Gen	Personal pronoun in genitive case, 3rd person singular	της, του, αυτού, αυτής, τού
	Pron-Pers-Sg3-Nom	Personal pronoun in nominative case, 3rd person singular	το, τη, αυτό, αυτή, αυτός

Umbrella tag	Tag	Description	Examples
	Pron-Rel	Relative pronoun invariant in case and number	που, ό,τι, ο,τι
	Pron-Rel-Pl-Acc	Relative pronoun in accusative case, plural	οποία, οποίες, οποιαδήποτε, όσα, όσους
	Pron-Rel-Pl-Gen	Relative pronoun in genitive case, plural	οποίων, όσων, οσωνδήποτε, οσωνδήποτε, ορωνδήποτε
	Pron-Rel-Pl-Nom	Relative pronoun in nominative case, plural	οποία, οποίες, οποιαδήποτε, όσα, οποίοι
	Pron-Rel-Sg-Acc	Relative pronoun in accusative case, singular	οποία, οποίο, όσο, οποιαδήποτε, όποια
	Pron-Rel-Sg-Gen	Relative pronoun in genitive case, singular	οποίας, οποίου, όποιου, οποιουδήποτε, όσης
	Pron-Rel-Sg-Nom	Relative pronoun in nominative case, singular	οποία, οποίο, όσο, οποίος, οποιαδήποτε
Punct	Punct	Punctuation mark	:"}-]
	Punct-Comma	Comma	,
	Punct-Sent	Punctuation mark, end of sentence	.!?;
V	V-Sg1	Verb, 1st person singular	θέλω, λέω, δω, έχω, πήρα
	V-Sg2	Verb, 2nd person singular	εξελίξεις, δηλώσεις, επιχειρήσεις, απαιτήσεις, κυβερνήσεις
	V-Sg3	Verb, 3rd person singular	έχει, πρέπει, είχε, μπορεί, φορά
	V-P11	Verb, 1st person plural	έχουμε, κάνουμε, είχαμε, μπορούμε, δούμε
	V-P12	Verb, 2nd person plural	εφαρμόζετε, έχετε, δείτε, μπορείτε, καλύψετε
	V-P13	Verb, 3rd person plural	έχουν, μπορούν, είχαν, υπάρχουν, αποτελούν
	V-Pass-Sg1	Passive verb, 1st person singular	είμαι, ήμουν, θυμάμαι, γεννηθώ, φοβηθώ

Umbrella tag	Tag	Description	Examples
	V-Pass-Sg2	Passive verb, 2nd person singular	είσαι, αισθάνεσαι, κουράζεσαι, προσαρμοστείς, βρίσκεσαι
	V-Pass-Sg3	Passive verb, 3rd person singular	είναι, ήταν, βρίσκεται, έγινε, γίνεται
	V-Pass-Pl1	Passive verb, 1st person plural	είμαστε, βρισκόμαστε, βρεθήκαμε, οδηγηθούμε, δεχθούμε
	V-Pass-Pl2	Passive verb, 2nd person plural	είστε, θυμηθείτε, απευθυνθείτε, οργανωθείτε, προμηθευτείτε
	V-Pass-Pl3	Passive verb, 3rd person plural	είναι, ήταν, βρίσκονται, λέγονται, γίνονται
	V-Imp-Pl	Imperative verb, plural	εφαρμόζετε, έχετε, θέλετε, μπορείτε, καλύψετε
	V-Imp-Sg	Imperative verb, singular	χώρα, δήλωσε, μέτρα, μήνα, άρχισε
	V-Imp-Pass-Pl	Imperative verb, passive, plural	θυμηθείτε, απευθυνθείτε, αποτοξινωθείτε, προμηθευτείτε, οργανωθείτε
	V-Imp-Pass-Sg	Imperative verb, passive, singular	ελέγχου, περιορίσου, άνελθε, έπερθε, έπελθε
	V-PrPart	Active participle	έχοντας, φέρνοντας, τονίζοντας, κρατώντας, κάνοντας
	V-PaPart-Pl-Acc	Passive participle in accusative case, plural	κείμενα, δεδομένα, αγαπημένες, εργαζόμενους, ορισμένες
	V-PaPart-Pl-Gen	Passive participle in genitive case, plural	αυξημένων, εγγεγραμμένων, αγαπημένων, περιγεγραμμένων, καλεσμένων
	V-PaPart-Pl-Nom	Passive participle in nominative case, plural	κείμενα, δεδομένα, συνηθισμένα, ορισμένες, αγαπημένες

Umbrella tag	Tag	Description	Examples
	V-PaPart-Pl-Voc	Passive participle in vocative case, plural	κείμενα, δεδομένα, συνηθισμένα, ορισμένες, αγαπημένες
	V-PaPart-Sg-Acc	Passive participle in accusative case, singular	δοσμένο, αυξημένη, οπλισμένο, φτιαγμένο
	V-PaPart-Sg-Gen	Passive participle in genitive case, singular	περιγεγραμμένου, εγγεγραμμένου, βραβευμένου, βασισμένης, περασμένης
	V-PaPart-Sg-Nom	Passive participle in nominative case, singular	δοσμένο, αυξημένη, οπλισμένο, γραμμένο
	V-PaPart-Sg-Voc	Passive participle in vocative case, singular	δοσμένο, αυξημένη, οπλισμένο, φτιαγμένο
	V-Inv	Infinitive verb	κάνει, αναφέρει, δώσει
	V-Inv-Pass	Infinitive verb, passive	γίνει, ψηφιστεί, ολοκληρωθεί

4.5.5.14 Hebrew Part-of-Speech Tagging

Shows the tag set available for defining Hebrew custom entities. The tag names are accompanied by a brief description and one or more examples. Hebrew words have a rich morphology. Words often consist of several morphemes which are separate words in other languages: prepositions, conjunctions, definiteness markers.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj-Pl	Plural adjective	משולבות, חיוורות
	Adj-Pl-Conj	Plural adjective with conjunction	ואדומים, ונוחים
	Adj-Pl-Conj-Prep	Plural adjective with conjunction and preposition	ושבחוכם
	Adj-Pl-Prep	Plural adjective with preposition	מהממולאים, לבעייתיות
	Adj-Sg	Singular adjective	מוחק, טוב, עצבנית
	Adj-Sg-Conj	Singular adjective with conjunction	וחשאית, ועירנית, וחמים

Umbrella Tag	Complete Tag	Description	Examples
	Adj-Sg-Conj-Prep	Singular adjective with conjunction and preposition	ובמקביל
	Adj-Sg-Prep	Singular adjective with preposition	בטוב, בצהוב, מרחוק
Adv	Adv	Adverb	שוב, מעולם, לא
	Adv-Conj	Adverb with conjunction	ואז, ולבסוף
	Adv-Conj-Prep	Adverb with conjunction and preposition	ומכאן
Agr	Agr	Agreement particle	יש, אין
	Agr-Pl	Agreement particle, plural	ישנם
	Agr-Sg	Agreement particle, singular	איני
At	At	Accusative marker	את
Com	Com	Complementizer	כאשר, אם
Conj	Conj	Coordinating conjunction	אבל, או
Det	Det-B	Determiner, both numbers	כל
	Det-B-Conj	Determiner, both numbers, with conjunction	וכל
	Det-B-Prep	Determiner, both numbers, with preposition	בכל
	Det-B-Conj-Prep	Determiner, both numbers, with conjunction and preposition	ומכל
	Det-H	Determiner, definiteness marker	ה
	Det-Pl	Determiner plural	אותם, אותן
	Det-Pl-Conj	Determiner plural with conjunction	ואותנו
	Det-Pl-Prep	Determiner plural with preposition	באותם

Umbrella Tag	Complete Tag	Description	Examples
	Det-Pl-Conj-Prep	Determiner plural with conjunction and preposition	ובאותן
	Det-Sg	Determiner singular	אותו, איזה
	Det-Sg-Conj	Determiner singular with conjunction	וחלק
	Det-Sg-Conj-Prep	Determiner singular with conjunction and preposition	ובאותה
	Det-Sg-Prep	Determiner singular with preposition	לאיזשהו
Ham	Ham	Yes/no question word	כי
Nn	Nn-Pl	Plural noun	ימים, חולים
	Nn-Pl-Conj	Plural noun with conjunction	ופסים
	Nn-Pl-Conj-Prep	Plural noun with conjunction and preposition	ובנעליים, ולחצרות
	Nn-Pl-Prep	Plural noun with preposition	בכתומים, לענייני
	Nn-Sg	Singular noun	סיפור, סקרנות
	Nn-Sg-Conj	Singular noun with conjunction	ואהבה, וגודל
	Nn-Sg-Conj-Prep	Singular noun with conjunction and preposition	ומאילן
	Nn-Sg-Prep	Singular noun with preposition	במרכזו, בצורה, במסיבת
Num	Num	Numeral	אחד, שני, 1975
	Num-Conj	Numeral with conjunction	ואחת, ושלושה
	Num-Prep	Numeral with preposition	לשני, בשלוש
	Num-Conj-Prep	Numeral with conjunction and preposition	ושמהעשר
Pos	Pos	Possessive item	של
Prep	Prep	Preposition	כדי, מעל, לצד

Umbrella Tag	Complete Tag	Description	Examples
	Prep-Conj	Preposition with conjunction	ועל, ומן
Pron	Pron-Pl1	Pronoun, first person plural	אנחנו
	Pron-Pl1-Conj	Pronoun, first person plural with conjunction	ואנחנו
	Pron-Pl1-Prep	Pronoun, first person plural with preposition	לנו
	Pron-Pl2	Pronoun, second person plural	אתם
	Pron-Pl2-Conj	Pronoun, second person plural with conjunction	ואתם
	Pron-Pl2-Prep	Pronoun, third person plural with preposition	מביניכם
	Pron-Pl3	Pronoun, third person plural	הם, הן
	Pron-Pl3-Conj	Pronoun, third person plural with conjunction	והם, ואלה
	Pron-Pl3-Prep	Pronoun, third person plural with preposition	להם, מהם, להן
	Pron-Sg1	Pronoun, first person singular	אני
	Pron-Sg1-Conj	Pronoun, first person singular with conjunction	ואני
	Pron-Sg1-Prep	Pronoun, first person singular with preposition	לידי
	Pron-Sg2	Pronoun, second person singular	את, אתה
	Pron-Sg2-Conj	Pronoun, second person singular with conjunction	ואת
	Pron-Sg2-Prep	Pronoun, second person singular with preposition	בגללך
Pron-Sg3	Pronoun, third person singular	היא, הוא	

Umbrella Tag	Complete Tag	Description	Examples
	Pron-Sg3-Conj	Pronoun, third person singular with conjunction	והיא, והוא
	Pron-Sg3-Prep	Pronoun, third person singular with preposition	לה, אליו, לך
Prop	Prop	Proper noun	איציק, ישראל
	Prop-Conj	Proper noun with conjunction	וריצ'רד, וקאן
	Prop-Conj-Prep	Proper noun with conjunction and preposition	ולאנטונינוס
	Prop-Prep	Proper noun singular with preposition	לקלייר, מירושלים
Punct	Punct	Punctuation mark	– ; : ' "
	Punct-Close	Closing bracket)
	Punct-Comma	Comma	,
	Punct-Open	Opening bracket	(
	Punct-Quote	Quote	"
	Punct-Sent	Sentence ending mark	. ! ?
Qw	Qw	Question/WH word	מה, איך, כמה
	Qw-Conj	Question/WH word with conjunction	ומה
	Qw-Prep	Question/WH word with preposition	כמי
V	V-Imp	Imperative verb	בואי, שימי
	V-Imp-Conj	Imperative verb with conjunction	ותברח, וצאו
	V-Inf	Infinitive verb	למנוע, להחביא
	V-Inf-Conj	Infinitive verb with conjunction	ולקבוע, וללמוד
	V-P1	Finite verb, plural	שותקים, ממשיכים, מגיעים
	V-P11	Finite verb, first person plural	נטיל, נודה, נשחק

Umbrella Tag	Complete Tag	Description	Examples
	V-P11-Conj	Finite verb, first person plural with conjunction	ונגיע
	V-P12	Finite verb, second person plural	התלבשתם, מצאתם
	V-P12-Conj	Finite verb, second person plural with conjunction	ותווכחו
	V-P13	Finite Verb, third person plural	הסיעו, עברו, הסירו
	V-P13-Conj	Finite Verb, third person plural with conjunction	ויהיו
	V-P1-Conj	Finite Verb, plural with conjunction	ונחשפות
	V-Sg	Finite Verb, singular	אומרת
	V-Sg1	Finite verb, first person singular	עשיתי, הייתי
	V-Sg1-Conj	Finite verb, first person singular with conjunction	והתפעלתי
	V-Sg2	Finite verb, second person singular	ידעת, היית
	V-Sg2-Conj	Finite verb, second person singular with conjunction	והיית
	V-Sg3	Finite verb, third person singular	הייתה, אמרה
	V-Sg3-Conj	Finite verb, third person singular with conjunction	ואמרה
	V-Sg-Conj	Finite verb, singular with conjunction	ונמערך

4.5.5.15 Hungarian Part-of-Speech Tagging

Shows the tag set available for defining Hungarian custom entities. The tag names are accompanied by a brief description and one or more examples

Umbrella tag	Complete Tag	Description	Examples
Adj	Adj-Inv	Adjective case and number invariant	úgynevezett, kis, ún
	Adj-Pl-Acc	Plural adjective in accusative case	kérőket
	Adj-Pl-Acc-Poss	Plural adjective in accusative case with possessive suffix	írországtatlanjukét, oginyíróldöcögőinket, jobaházatlanai-két
	Adj-Pl-Case	Plural adjective in case other than Nom and Acc	Ezeknek, azoknál, ezekből
	Adj-Pl-Case-Poss	Plural adjective in case other than Nom and Acc with possessive suffix	quattrocentookosukon, bú-bánatvölgyiővónkén, úrhidatlanbaúzószűkre
	Adj-Pl-Nom	Plural adjective in nominative case	azok, magyarok, helyiek
	Adj-Pl-Nom-Poss	Plural adjective in nominative case with possessive suffix	úricsiraelőttikorink, thesaurusájtatosatok, sőregpusztánbőkésűik
	Adj-Sg-Acc	Singular adjective in accusative case	azt, ezt
	Adj-Sg-Acc-Poss	Singular adjective in accusative case with possessive suffix	eltűnését
	Adj-Sg-Case	Singular adjective in case other than Nom and Acc	újra, azon, abba, román
	Adj-Sg-Case-Poss	Singular adjective in case other than Nom and Acc with possessive suffix	igazán, nagyjából, mélyén
	Adj-Sg-Nom	Singular adjective in nominative case	az, ez, orosz, olyan
	Adj-Sg-Nom-Poss	Singular adjective in nominative case with possessive suffix	utasa, fegyveresei, nagyménője

Umbrella tag	Complete Tag	Description	Examples
Adv	Adv	Adverb	is, csak, nem, sem
Conj	Conj	Conjunction	és, hogy, de, meg
Det	Det	Determiner	a, egy, egy-egy
Interj	Interj	Interjectuin	ne, e, Na, na, Ne, Ja, ugye, E, Ugye, no
Nn	Nn-Inv	Noun case and number invariant	NDK-e, tsz-e, USA-i
	Nn-Inv-Nom	Noun in nominative case, number invariant	Ceaușescu, Fidesz-KDNP, B7-es
	Nn-Inv-Acc	Noun in accusative case, number invariant	detiizúzúvépiasztert, ördögöspiasztert
	Nn-Inv-Case	Noun in case other than Nom and Acc, number invariant	NDK-del, FÁK-gal, UV-hel
	Nn-Pl-Acc	Plural noun in accusative case	azokat, magukat, akiket
	Nn-Pl-Acc-Poss	Plural noun in accusative case with possessive suffix	megszűnésüket, tőkéjüket, rezdülésünket
	Nn-Pl-Case	Plural noun case other than Nom and Acc	mértékben, területeken, akiknek
	Nn-Pl-Case-Poss	Plural noun in case other than Nom and Acc with possessive suffix	életükből, ígéreteikhez, felszámolásukról
	Nn-Pl-Nom	Plural noun in nominative case	akik, amelyek, kérdések
	Nn-Pl-Nom-Poss	Plural noun in nominative case with possessive suffix	százalékuk, többségük, nehézfegyvereik
	Nn-Sg-Acc	Singular noun in accusative case	amit, azt, megállapodást
	Nn-Sg-Acc-Poss	Singular noun in accusative case with possessive suffix	százalékát, napirendjét, tárgyalását
Nn-Sg-Case	Singular noun in case other than Nom and Acc	annak, azzal, abban	

Umbrella tag	Complete Tag	Description	Examples
	Nn-Sg-Case-Poss	Singular noun in case other than Nom and Acc with possessive suffix	számára, alapján, esetében
	Nn-Sg-Nom	Singular noun in nominative case	nem, ország, minden
	Nn-Sg-Nom-Poss	Singular noun in nominative case with possessive suffix	része, aránya, célja
Num	Num	Numeral	3, 100, 24
	Num-Ord	Ordinal numeral	63., 17., 2013.
	Num-Ord-Pl-Acc	Ordinal numeral in accusative case, plural	úkalattiakon-elsőkét, Nyúllágyőri-elsőkét, múmiaalattim-elsőkét
	Num-Ord-Pl-Case	Ordinal numeral in case other than Nom and Acc, plural	gülbabaödémaelsőkig, jet-tácskaim-elsőkbe, úrtalanhozég-elsőkbe
	Num-Ord-Pl-Nom	Ordinal numeral in nominative case, plural	Máhomfa-tarpai-elsők, zúdulgass-Luda-elsők, ihasdvajkosáé-elsők
	Num-Ord-Sg-Acc	Ordinal numeral in accusative case, singular	pneumatikcókómokelsőt, oderásómegaege-elsőt, möcinyiberídd-elsőt
	Num-Ord-Sg-Case	Ordinal numeral in case other than Nom and Acc, singular	harmadiknak, elsőként, negyedévhez
	Num-Ord-Sg-Nom	Ordinal numeral in nominative case, singular	első, második, harmadik, tizenhetedik
	Num-Pl-Acc	Numeral in accusative case, plural	óhutaiegével-sokakat, gősfátlanának-ötökét, xéniastop-pere-ötökét
	Num-Pl-Case	Numeral in case other than Nom and Acc, plural	egyikünknek
	Num-Pl-Nom	Numeral in nominative case, plural	többek, sokak, ellenfelek
	Num-Sg-Acc	Numeral in accusative case, singular	sokat, többet, annyit

Umbrella tag	Complete Tag	Description	Examples
	Num-Sg-Case	Numeral in case other than Nom and Acc, singular	egyre, mennyire, egyben
	Num-Sg-Nom	Numeral in nominative case, singular	két, több, milliárd
Postp	Postp	Postposition	szerint, után, között, előtt
Pref	Pref	Prefix	el, ki, be, viszont
Pron	Pron-Pers-Pl-Acc	Personal pronoun in accusative case, plural	őket, bennünket
	Pron-Pers-Pl-Case	Personal pronoun in case other than Nom and Acc, plural	rájuk, velük, nekünk
	Pron-Pers-Pl-Nom	Personal pronoun in nominative case, plural	mi, ők, Mi, Ők, maguk
	Pron-Pers-Sg-Acc	Personal pronoun in accusative case, singular	őt, magát, engem
	Pron-Pers-Sg-Case	Personal pronoun in case other than Nom and Acc, singular	hozzá, rá, benne, vele
	Pron-Pers-Sg-Nom	Personal pronoun in nominative case, singular	ő, én, maga, te
Punct	Punct	Punctuation mark	: -) (
	Punct-Comma	Comma	,
	Punct-Sent	End of sentence mark	. ? ; !
V	V-AdvPart	Adverbial participle	illetve, nyitva, éve
	V-FutPart-Pl-Acc	Future participle in accusative case, plural	nevefajta-oldandókét, vékenyetlen-eendőkét
	V-FutPart-Pl-Case	Future participle in case other than Nom and Acc, plural	gügyöréssze-eendőkbe, ggyarássza-eendőkke, xenszám-óivandókén
	V-FutPart-Pl-Nom	Future participle in nominative case, plural	nukleonlárvaérzendők, Kínár-aógandókázandók, Lucócskákrolömlendők

Umbrella tag	Complete Tag	Description	Examples
	V-FutPart-Sg-Acc	Future participle in accusative case, plural	Adolárnál-sés-eendőt, Írorszáigig-gátlandót, tétessünk-óztatandót
	V-FutPart-Sg-Case	Future participle in case other than Nom and Acc, plural	jonásmókárólömlendőn, bájukbeliekéi-eendőn, jerünk-xéniásítandón
	V-FutPart-Sg-Nom	Future participle in nominative case, plural	követendő
	V-Inf	Infinitive	tudni, lenni, mondani, adni
	V-PaPart-Inv-Acc	Past participle in accusative case, invariant in number	wesselényinyireittát, únyinómenklatúráztát, ivargímszer-úskódtét
	V-PaPart-Inv-Case	Past participle in case other than Nom and Acc, invariant in number	következtében, ígértével, mentén
	V-PaPart-Inv-Nom	Past participle in accusative case, invariant in number	mondta, látta, közölte
	V-PaPart-Pl-Acc	Past participle in accusative case, plural	menekülteket, törtéteket
	V-PaPart-Pl-Case	Past participle in case other than Nom and Acc, plural	menekültekről, törtétekkal, törtétekről
	V-PaPart-Pl-Nom	Past participle in nominative case, plural	menekültek, vittek, kerültek
	V-PaPart-Sg-Acc	Past participle in accusative case, plural	folytat
	V-PaPart-Sg-Case	Past participle in case other than Nom and Acc, plural	mentén, tapasztalhoz, akkreditáltként
	V-PaPart-Sg-Nom	Past participle in nominative case, plural	tartott, adott, kialakított
	V-P11	Finite verb, 1st person plural	tudunk, tartunk, kapunk
	V-P12	Finite verb, 2nd person plural	törtétek, érkeztek, igyekeztek
	V-P13	Finite verb, 3rd person plural	vannak, voltak, lesznek

Umbrella tag	Complete Tag	Description	Examples
	V-PrPart-Pl-Acc	Present participle in accusative case, plural	szervezőket, hadisírgondozókat
	V-PrPart-Pl-Case	Present participle in case other than Nom and Acc, plural	vezetőkkel, szereplőknek, alapítóknak
	V-PrPart-Pl-Nom	Present participle in nominative case, plural	szereplők, képviselők, beszélők
	V-PrPart-Sg-Acc	Present participle in accusative case, plural	illetőt, adót, tényezőt
	V-PrPart-Sg-Case	Present participle in case other than Nom and Acc, plural	egyelőre, jövőben, sajtótájékoztatón
	V-PrPart-Sg-Nom	Present participle in nominative case, plural	hasonló, álló, lévő
	V-Sg1	Finite verb, 1st person plural	vagyok, voltam, érzem
	V-Sg2	Finite verb, 2nd person plural	vagy, ennél, lehetsz
	V-Sg3	Finite verb, 3rd person plural	volt, van, kell, lehet

4.5.5.16 Indonesian Part-of-Speech Tagging

Shows the tag set available for defining Indonesian custom entities. The tag names are accompanied by a brief description and one or more examples.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj	Adjective excluding superlatives	besar
	Adj-Sup	Superlative	terbesar terpandai
Adv	Adv	Adverb	pelan-pelan
Aux	Aux	Auxiliary verb	perlu
Conj	Conj	Conjunction	atau

Umbrella Tag	Complete Tag	Description	Examples
Cop	Cop	Copula	adalah
Date	Date	Date, including month, day	20-06-2014
Det	Det	Determiner	beberapa
Func	Func	Other function, and miscellaneous word	dst
Interj	Interj	Interjection	oi
Neg	Neg	Negation verb	takkan belum
Nn	Nn	Default noun	buku
	Nn-Loc	Locative nominal	tepi
	Nn-Net	Internet or e-mail address	www.sap.com
	Nn-Pl	Plural noun	buku-buku meja-meja
Num	Num	Numeral, written and digits incl. roman, ordinal, and time	28
Part	Part	Particle verb	pun
Prep	Prep	Preposition	di
Pron	Pron	Other pronoun	lainnya
	Pron-Pers	Personal pronoun	saya
	Pron-Yang	Relative pronoun	yang
Prop	Prop	Proper noun	Soeharto
Punct	Punct	Other punctuation	::
	Punct-Comma	Comma	,
	Punct-Sent	Sentence-ending punctuation	!. ?
	Punct-Open	Opening punctuation	([< {
	Punct-Close	Closing punctuation)] > }

Umbrella Tag	Complete Tag	Description	Examples
	Punct-Quote	Single or double quote	"
Ques	Ques	Question word	mengapa
Verb	Verb	Other verb not covered by other tags	kirim
	Verb-Pass	Passive verb	diberikan
	Verb-Pp	Past participle verb	terbuka

4.5.5.17 Italian Part-of-Speech Tagging

Shows the tag set available for defining Italian custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj-ex	The prefix ex	ex
	Adj-Pl	Plural adjective (includes ordinals: secondi)	belle
	Adj-PrPart-Pl	Plural present participle verb	meditanti, destreggianti
	Adj-PrPart-Pl-Pron	Plural present participle verb with attached clitic	fasciantemela, quietanteti
	Adj-PrPart-Sg	Singular present participle verb	meditante, destreggiante
	Adj-PrPart-Sg-Pron	Singular present participle verb with attached clitic	epurantelo, andantevi
	Adj-Sg	Singular adjective (includes ordinals: secondo, 2°)	buono, narcisistico
Adv	Adv	Adverb	fumettisticamente
Aux	Aux	Finite auxiliary (be and have)	saranno, avrete
	Aux-Ger	Gerundive auxiliary	essendo, avendo
	Aux-Impv	Imperative auxiliary	sii, abbi

Umbrella Tag	Complete Tag	Description	Examples
	Aux-Inf	Infinitive auxiliary	esser, aver
	Aux-PaPart-Pl	Plural past participle auxiliary	avuti, avute
	Aux-PaPart-Sg	Singular past participle auxiliary	avuta, avuto
	Aux-PrPart-Pl	Plural present participle auxiliary	essenti, aventi
	Aux-PrPart-Sg	Singular present participle auxiliary	essente, avete
Conj	Conj	Conjunction	tuttavia
	Conj-Adv	Interrogative adverb	quando, dove, come
	Conj-che	The connector che	ch', che
	Conj-Coord	Coordinating conjunction	ed, e/o
	Conj-Pre	First part of a multiword conjunction	dato che
Det/Pron	Det/Pron-Int-Pl	Plural interrogative determiner or pronoun	quanti soldi
	Det/Pron-Int-Sg	Singular interrogative determiner or pronoun	qual
	Det/Pron-Poss-Pl	Plural possessive determiner or pronoun	mie, vostri
	Det/Pron-Poss-Sg	Singular possessive determiner or pronoun	nostro, sua
	Det/Pron-Quant	Invariant quantifying determiner or pronoun	qualunque, qualsivoglia
	Det/Pron-Quant-Pl	Plural quantifying determiner or pronoun	molti uomini
	Det/Pron-Quant-Sg	Singular quantifying determiner or pronoun	molta gente
Det	Det-Pl	Plural determiner	quei
	Det-Pre	Pre-determiner	tutto il giorno

Umbrella Tag	Complete Tag	Description	Examples
	Det-Sg	Singular determiner	quel
Interj	Interj	Interjection or onomatopoeia	uhi, perdiana, eh
Nn	Nn-Letter	Lowercase and uppercase letters, by themselves or followed by a period or right parenthesis	b, N
	Nn-Net	URL, e-mail address	www.inxight.com, info@inxight.com
	Nn-Pl	Plural noun	case
	Nn-Sg	Singular noun	casa, balsamo
Num	Num	Numeric expression (in digits)	+5, 23.05, 3,45, 1997
Prep	Prep	Preposition	tra, con
	Prep-a	Preposition a	a
	Prep-da	Preposition da	da
	Prep-Det-Pl	Combination preposition and plural determiner	sulle, sugl', pegli
	Prep-Det-Pl-a	Combination a and plural determiner	ai, alle
	Prep-Det-Pl-da	Combination da and plural determiner	dalle
	Prep-Det-Pl-di	Combination di and plural determiner	delle
	Prep-Det-Sg	Combination preposition and singular determiner	sullo, nella
	Prep-Det-Sg-a	Combination a and singular determiner	al, allo
	Prep-Det-Sg-da	Combination da and singular determiner	dalla
Prep-Det-Sg-di	Combination di and singular determiner	della	

Umbrella Tag	Complete Tag	Description	Examples
	Prep-di	Preposition di	di
	Prep-Pre	First word of a multiword preposition	per mezzo
Pron	Pron	Invariant pronoun	sé
	Pron-chi	Interrogative pronoun chi	chi
	Pron-Clitic	Clitic pronoun	vi, ne, mi, glielo
	Pron-Clitic-Pre	First of a two-clitic sequence	ce, ve
	Pron-Indef-Pl	Plural indefinite pronoun	Tutti amano le vacanze.
	Pron-Indef-Sg	Singular indefinite pronoun	qualcuno
	Pron-Pl	Plural pronoun	noi
	Pron-Rel	Invariant relative pronoun	cui
	Pron-Rel-Pl	Plural relative pronoun	i bambini i quali
	Pron-Rel-Sg	Singular relative pronoun	il bambino il quale
	Pron-Sg	Singular pronoun	lei, lui
Prop	Prop	Proper noun	Bernardo, Monte Isola
Punct	Punct	Punctuation	: - \
	Punct-Comma	Comma	,
	Punct-Sent	Sentence punctuation	. ! ? ;
V/Adj	V/Adj-PaPart-Pl	Plural past participle verb or adjective	riposti, offuscate
	V/Adj-PaPart-Pl-Pron	Plural past participle verb or adjective, with attached clitic	telatesele, assestatici
	V/Adj-PaPart-Sg	Singular past participle verb or adjective	sbudellata
	V/Adj-PaPart-Sg-Pron	Singular past participle verb or adjective, with attached clitic	commossosi, ingranditomi
V	V-Fin	Finite verb	blatereremo

Umbrella Tag	Complete Tag	Description	Examples
	V-Fin-Pron	Finite verb with attached clitic	trattansi, leggevansi
	V-Ger	Gerund	adducendo, intervistando
	V-Ger-Pron	Gerund with attached clitic	saziandotele, appurandolo
	V-Impv	Imperative verb	Va' a casa!
	V-Impv-Pron	Imperative verb with attached clitic	russateli, planaci
	V-Inf	Infinitive verb	sciupare, trascinar
	V-Inf-Pron	Infinitive verb with attached clitic	spulciarsi, risucchiarsi

4.5.5.18 Japanese Part-of-Speech Tagging

Shows the tag set available for defining Japanese custom entities. The tag names are accompanied by a brief description and one or more examples.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj	Adjective	赤い、大きい
	Adj-D	Adjective denoting modality	(て)ほしい、(て)よい
Adnom	Adnom	Prenominal nominal	この、そんな
Adv	Adv	Adverb	ゆっくり、じっと
Aux	Aux	Auxiliary verb	だ、です、ない
Case	Case	Case marker	が、を、さえ
Conj	Conj	Conjunction	そして、しかし
Interj	Interj	Interjection	さあ、えっ
Nn	Nn	Noun	先生、分析、ファイル
	Nn-Adv	Noun commonly used adverbially	今日、午後、1月

Umbrella Tag	Complete Tag	Description	Examples
	Nn-Ascii	ASCII character, sequence or word	computer
	Nn-D	Formal noun; non-contentful noun	こと、の、もの、よう
	Nn-Pron	Pronoun	あなた、私、ここ
	Nn-Prop	Proper noun	山田、富士山
Num	Num	Numeric nominal	2000年、95%
Pre	Pre	Nominal prefix	お(水)、高(品質)
Punct	Punct	Punctuation	" : # @
	Punct-Close	Closing punctuation)、}、」
	Punct-Comma	Comma	、
	Punct-Open	Opening punctuation	(、{、「
	Punct-Sent	Sentence-ending punctuation	。 ?
Suf	Suf	Suffix	さん
Verb	Verb	Verb	読む、理解する
	Verb-D	Verb denoting modality and aspect	はじめる、できる、(て)いる

4.5.5.19 Korean Part-of-Speech Tagging

Shows the tag set available for defining Korean custom entities. The tag names are accompanied by a brief description and one or more examples.

Umbrella Tag	Complete Tag	Description	Examples
Adv	Adv	Adverb	만일, 그러면
Case	Case	Case marker	을, 를
Det	Det	Determiner	이, 그

Umbrella Tag	Complete Tag	Description	Examples
Interj	Interj	Interjection	아이고, 어머
Nn	Nn	Noun	책, 코끼리
	Nn-Ascii	ASCII characters, noun	copyright, Web
	Nn-Case	Noun case marked	호박은
	Nn-Case-Acc	Noun case marked accusative	조각을
	Nn-Case-Conj	Conjoined case marked noun	고양이와는, 토끼와는
	Nn-Case-Conj-Pl	Conjoined case marked plural noun	교인들하고는
	Nn-Case-Disj	Disjunctive case marked noun	여기까지나
	Nn-Case-Disj-Pl	Disjunctive case marked plural noun	박사들까지나
	Nn-Case-Gen	Noun case marked genitive	인간의
	Nn-Case-Pl	Noun case marked plural	선생님들은, 군인들은
	Nn-Case-Pl-Acc	Noun case marked plural accusative	학생들을
	Nn-Case-Pl-Gen	Noun case marked plural genitive	교수들의
	Nn-Conj	Conjoined noun	강아지와, 사자와
	Nn-Conj-Pl	Conjoined plural noun	교인들하고, 친구들하고
	Nn-Disj	Disjunctive noun	짐승이나, 과학자나
	Nn-Disj-Pl	Disjunctive plural noun	약사들이나, 화가들이나
	Nn-Gen	Genitive noun	참패의, 기자의
	Nn-Gen-Pl	Genitive plural noun	주자들의, 정치인들의
	Nn-Pl	Noun plural	사람들, 박사들
	Num	Num	Numeric expression

Umbrella Tag	Complete Tag	Description	Examples
	Num-Abl	Numeric expression (ablative)	3학년부터, 1987부터
	Num-Acc	Numeric expression (accusative)	10명을, 560을
	Num-Conj	Numeric expression (conjunction)	3루타와
	Num-Dir	Numeric expression (directive)	1,314.40으로, 49.0으로
	Num-Disj	Numeric expression (disjunctive)	2배나, 10차레나
	Num-Gen	Numeric expression (genitive)	12대의, 890만주의
	Num-Goal	Numeric expression (goal)	사장까지, 194.5까지
	Num-Inc	Numeric expression (inclusive)	1승도, 1개사도, 19.경기도
	Num-Inst	Numeric expression (instrumental)	4번타자로, 3점포로
	Num-Lmt	Numeric expression (limitive)	13만, 4면만에
	Num-Loc	Numeric expression (locative)	113장에
	Num-Neg	Numeric expression (negative)	이밖에
	Num-Nom	Numeric expression (nominative)	12명이, 4위인
	Num-Src	Numeric expression (source)	5경기에서, 구단에서
	Num-Top	Numeric expression (topic)	3국은, 4조는, 1천800명은
	Num-Day	Numeric day expression	9일, 주5일
	Num-Meas	Numeric measure expression	10.20리터, 100야드
	Num-Money	Numeric currency expression	구원, 10달러로, 2달러대가
	Num-Month	Numeric month expression	1월초, 1월말

Umbrella Tag	Complete Tag	Description	Examples
	Num-Percent	Numeric percent expression	9%, 57%가, 7%까지
	Num-Time	Numeric time expression	10시, 3시간
	Num-Year	Numeric year expression	2000년, 95년
Pron	Pron	Pronoun	나, 너
	Pron-Case	Case marked pronoun	나는
	Pron-Case-Acc	Pronoun case marked accusative	너를
	Pron-Case-Conj	Pronoun case marked conjunctive	자네하고는
	Pron-Case-Conj-Pl	Pronoun case marked conjunctive plural	우리들하고는
	Pron-Case-Disj	Pronoun case marked disjunctive	자기나만은
	Pron-Case-Disj-Pl	Pronoun case marked disjunctive plural	저이들까지나
	Pron-Case-Gen	Pronoun case marked genitive	나의
	Pron-Case-Pl	Pronoun case marked plural	우리들은
	Pron-Case-Pl-Acc	Pronoun case marked plural accusative	우리들을
	Pron-Case-Pl-Gen	Pronoun case marked plural genitive	우리들의
	Pron-Conj	Conjoined pronoun	자네하고
	Pron-Conj-Pl	Conjoined plural pronoun	우리들하고
	Pron-Disj	Disjunctive pronoun	자기나
	Pron-Disj-Pl	Disjunctive plural pronoun	우리들이나
Pron-Pl	Plural pronoun	우리들, 그들	
Prop	Prop	Proper name	삼성전자, 서울대학교
	Prop-Case	Case marked proper name	현대건설은

Umbrella Tag	Complete Tag	Description	Examples
	Prop-Case-Acc	Proper name case marked accusative	고려대학교를
	Prop-Case-Conj	Proper name case marked conjunctive	나이키와는
	Prop-Case-Conj-Pl	Proper name case marked conjunctive plural	대표들과도
	Prop-Case-Disj	Proper name case marked disjunctive	소니에게나
	Prop-Case-Gen	Proper name case marked genitive	한국은행의
	Prop-Case-Pl	Proper name case marked plural	조사관들이, 변호사들은
	Prop-Case-Pl-Acc	Proper name case marked accusative plural	대표들을, 국무위원들을
	Prop-Conj	Conjoined proper name	동국제강과
	Prop-Conj-Pl	Conjoined plural proper name	장관들과
	Prop-Disj	Disjunctive proper name	UBS나
	Prop-Gen	Proper name case marked genitive	공정거래위원회의, 일본의
	Prop-Gen-Pl	Proper name case marked genitive plural	의원들의, 감사위원들의
	Prop-Pl	Proper name plural	변호사들, 의원들
Punct	Punct	Punctuation	;, '
	Punct-Comma	Punctuation comma	,
	Punct-Sent	Punctuation sentence	.
V	V-Fut	Future tense verb	판매하겠다, 시작하겠다
	V-Past	Past tense verb	출발했다, 몰랐었다
	V-PreMod	Pre-modifying verb	좋은
	V-Pres	Present tense verb	상회하다, 번거롭다

4.5.5.20 Neutral Part-of-Speech Tagging

Shows the tag set available for defining Neutral custom entities. The tag names are accompanied by a brief description and one or more examples.

Umbrella Tag	Complete Tag	Description	Examples
Nn	Nn-Net	/URL, e-mail address	www.sap.com
Num	Num	Number	300
Punct	Punct	Any punctuation that is not sentence ending	, ;
	Punct-Sent	Sentence-ending punctuation	. ! ?
Unknown	Unknown	Any token that is not Nn-Net, Num, Punct, or Punct-Sent	cat ভা.জসভা Đông

4.5.5.21 Norwegian (Bokmål) Part-of-Speech Tagging

Shows the tag set available for defining Bokmål custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	adr., ibid.
Adj	Adj-Comp	Comparative adjective	mindre, eldre
	Adj-Comp-Gen	Genitive comparative adjective	eldres
	Adj-Def-Pl	Definite and/or plural adjective	gretne, mørke, kalde
	Adj-Def-Pl-Gen	Genitive definite and/or plural adjective	fremmedes, uvørnes
	Adj-Def-Sg	Definite singular adjective	lille
	Adj-Def-Sg-Gen	Genitive definite singular adjective	lilles
	Adj-Indef-Sg	Indefinite singular adjective	skriftlig

Umbrella Tag	Complete Tag	Description	Examples
	Adj-Indef-Sg-Gen	Genitive indefinite singular adjective	skriftligs
	Adj-PaPart-Def-Pl	Definite and/or plural adjectival past participle	besøkte, befestede, dempede
	Adj-PaPart-Def-Pl-Gen	Genitive, definite and/or plural, adjectival past participle	besøktes, befestedes, dempedes
	Adj-PaPart-Indef-Sg	Adjectival past participle, indefinite singular	for kryptert kommunikasjon
	Adj-Pl	Plural adjective	små problemer
	Adj-Pl-Gen	Genitive plural adjective	smås
	Adj-PrPart	Adjectival present participle	begynnende, stirrende
	Adj-PrPart-Gen	Genitive adjectival present participle	reisendes, lekendes, gråtendes
	Adj-Sup	Superlative adjective	best, raskest, størst
	Adj-Sup-Def	Definite superlative adjective	fineste, innerste, viktigste
	Adj-Sup-Def-Gen	Genitive definite superlative adjective	finestes, innerstes, viktigstes
Adv	Adv	Adverb	ikke, oppe, heller
	Adv-Comp	Comparative adverb	lenger
	Adv-Sup	Superlative adverb	helst
Aux	Aux/V-Impv	Imperative auxiliary or main verb	bli
	Aux/V-Inf	Infinitive auxiliary or main verb	vaere
	Aux/V-Inf-SForm	S-Form of infinitive auxiliary or main verb	has
	Aux/V-PaPart	Past participle of auxiliary or main verb	hatt, vaert, blitt, fått
	Aux/V-Past	Past tense auxiliary or main verb	hadde, var, ble

Umbrella Tag	Complete Tag	Description	Examples
	Aux/V-Pres	Present tense auxiliary or main verb	har, er, blir, får
	Aux/V-Pres-SForm	S-Form of present tense auxiliary or main verb	has, fås
	Aux/V-PrPart	Present participle of auxiliary or main verb	havende, blivende
	Aux-Inf	Infinitive auxiliary verb	kunne, måtte
	Aux-Inf-SForm	S-Form of infinitive auxiliary verb	kunnes, måtte
	Aux-PaPart	Past participle of auxiliary verb	kunnet, måttet
	Aux-Past	Past tense auxiliary verb	kunne, måtte
	Aux-Pres	Present tense auxiliary verb	kan, vil
	Aux-Pres-SForm	S-Form of present tense auxiliary verb	villes, skulles
	Aux-PrPart	Present participle of auxiliary verb	villende, skullende
Cmpd	Cmpd-Part	Left compound part	kontor - og forretningsbygg
Conj	Conj	Subordinating or relativizing conjunction	som, mens
	Conj-Coord	Coordinating conjunction	og, eller
Det	Det-Art-Def-Pl	Definite plural determiner	disse
	Det-Art-Def-Sg	Definite singular determiner	på denne grunn
	Det-Art-Indef	Indefinite singular determiner	en, et
	Det/Pron-Comp	Determiner or pronoun, comparative	mer
	Det/Pron-Pl	Determiner or pronoun, plural	noen, hvilke, alle, andre
	Det/Pron-Pl-Gen	Determiner or pronoun, plural genitive	noens, alles, andres

Umbrella Tag	Complete Tag	Description	Examples
	Det/Pron-Sg	Determiner or pronoun, singular	hver, litt, alt
	Det/Pron-Sg-Gen	Determiner or pronoun, singular genitive	enhvers, annens
	Det/Pron-Sup	Determiner or pronoun, superlative	mest
	Det/Pron-Sup-Def	Determiner or pronoun, definite superlative	meste
Interj	Interj	Interjection	ja, herregud
Nn	Nn-Def-Pl	Definite plural noun	dørene, armene
	Nn-Def-Pl-Gen	Genitive definite plural noun	salongenes, kollegenes
	Nn-Def-Sg	Definite singular noun	flyet, klokken
	Nn-Def-Sg-Gen	Genitive definite singular noun	selskapets, spisestuens
	Nn-Indef-Pl	Indefinite plural noun	plasser, mapper
	Nn-Indef-Pl-Gen	Genitive indefinite plural noun	tiders, menneskers
	Nn-Indef-Sg	Indefinite singular noun	stol, stripe
	Nn-Indef-Sg-Gen	Genitive indefinite singular noun	topps
	Nn-Indef-SP	Indefinite singular or plural noun	lys, skritt
	Nn-Indef-SP-Gen	Genitive indefinite singular or plural noun	slags, lands, års
	Nn-Letter	Lowercase and uppercase letters	b, N
	Nn-Net	URL and e-mail address	www.inxight.com info@inxight.com
Num	Num	Cardinal numeric expression or plural cardinal number (spelled out)	-294, 4,6%, xii, 1.100 to, tre, fire

Umbrella Tag	Complete Tag	Description	Examples
	Num-Def-Sg	The number "one", definite singular (spelled out)	ene
	Num-Indef-Sg	The number "one", indefinite singular (spelled out)	en, ett
Ord	Ord	Ordinal number (in digits or spelled out)	7., første
Part	Part-Inf	Infinitival particle	å beskrive
Prep	Prep	Preposition	med, ut
	Prep-av	Preposition av	av
	Prep-for	Preposition for	for
	Prep-fra	Preposition fra	fra
	Prep-i	Preposition i	i
	Prep-paa	Preposition på	på bakgrunn
	Prep-ved	Preposition ved	ved
	Pron	Pron-Acc	Accusative pronoun
Pron-Nom		Nominative pronoun	han, hun
Pron-Poss-Pl		Possessive pronoun with plural agreement	sine
Pron-Poss-Sg		Possessive pronoun with singular agreement	sin
Prop	Prop	Proper name	Oslo, Arne
	Prop-Gen	Genitive proper name	Akers
Punct	Punct	Miscellaneous punctuation	- [
	Punct-Comma	Comma	,
	Punct-Quote	Quotation marks	" ' " "
	Punct-Sent	Sentence boundary punctuation	... ? ; !
V	V-Impv	Imperative verb	se, ta

Umbrella Tag	Complete Tag	Description	Examples
	V-Inf	Infinitive verb	komme, gjøre
	V-Inf-SForm	S-Form of infinitive verb	kan belastes
	V-PaPart	Past participle verb	reist, utpekt, stanset
	V-PaPart-SForm	S-Form of past participle verb	trivdes
	V-Past	Past tense verb	sa, vokste
	V-Past-SForm	S-Form of past tense verb	levdes, mistrivdes
	V-Pres	Present tense verb	vet, gir
	V-Pres-SForm	S-Form of present tense verb	flyttes, møtes, finnes, synes
	V-PrPart	Present participle verb	være avtakende

4.5.5.22 Norwegian (Nynorsk) Part-of-Speech Tagging

Shows the tag set available for defining Nynorsk custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	red.
Adj	Adj-Comp	Comparative adjective	mindre, eldre
	Adj-Def-Pl	Definite and/or plural adjective	gretne, mørke, kalde
	Adj-Def-Pl-Gen	Genitive definite plural adjective	framandes
	Adj-Def-Sg	Definite singular adjective	lisle, vesle
	Adj-Indef-Sg	Indefinite singular adjective	norsk
	Adj-Indef-Sg-Gen	Genitive indefinite singular adjective	nærleggjandes

Umbrella Tag	Complete Tag	Description	Examples
	Adj-PaPart-Def-Pl	Definite and/or plural adjective derived from past participle	kjende
	Adj-PaPart-Indef-Sg	Indefinite singular adjective derived from past participle	reist
	Adj-Pl	Plural adjective	ørsmåe
	Adj-PrPart	Adjective derived from present participle	begynnande, stirande
	Adj-PrPart-Gen	Genitive adjective derived from present participle	reisandes, leikandes
	Adj-Sup	Superlative adjective	best, raskast, størst
	Adj-Sup-Def	Definite superlative adjective	finaste, inste, viktigaste
Adv	Adv	Adverb	ikkje, no
	Adv-Comp	Comparative adverb	lenger, heller
	Adv-Sup	Superlative adverb	verst
Aux/V	Aux/V-Impv	Imperative auxiliary or main verb	ver
	Aux/V-Inf	Infinitive auxiliary or main verb	bli
	Aux/V-Inf-SForm	S-Form of infinitive auxiliary or main verb	havast, fåast
	Aux/V-PaPart	Past participle of auxiliary or main verb	hatt, vore, blitt, fått
	Aux/V-Past	Past tense auxiliary or main verb	hadde, var, blei, fekk
	Aux/V-Pres	Present tense auxiliary or main verb	har, er, blir, får
	Aux/V-Pres-SForm	S-Form of present tense auxiliary or main verb	verast, havast, lytast
	Aux/V-PrPart	Present participle of auxiliary or main verb	havande, blivande, verande

Umbrella Tag	Complete Tag	Description	Examples
Aux	Aux-Inf	Infinitive auxiliary verb	kunne, måtte
	Aux-PaPart	Past participle of auxiliary verb	vilja, måtta
	Aux-Past	Past tense auxiliary verb	kunne, måtte
	Aux-Pres	Present tense auxiliary verb	kan, vil
	Aux-PrPart	Present participle of auxiliary verb	viljande, kunnande
Cmpd	Cmpd-Part	Left compound part	kontor - og forretningsbygg
Conj	Conj	Subordinating or relativizing conjunction	som, mens
	Conj-Coord	Coordinating conjunction	og, eller
Det/Pron	Det/Pron-Comp	Comparative determiner or pronoun	meir
	Det/Pron-Pl	Plural determiner or pronoun	alle
	Det/Pron-Sg	Singular determiner or pronoun	nokon
	Det/Pron-Sup	Superlative determiner or pronoun	mest
	Det/Pron-Sup-Def	Definite superlative determiner or pronoun	meste
Det	Det-Art-Def-Pl	Definite plural determiner (article or demonstrative pronoun)	dei, desse
	Det-Art-Def-Sg	Definite singular determiner (article or demonstrative pronoun)	denne artikkelen
	Det-Art-Indef	Indefinite singular determiner	eit
Interj	Interj	Interjection	hei, tjo
Nn	Nn-Def-Pl	Definite plural noun	dørene, armane
	Nn-Def-Pl-Gen	Genitive definite plural noun	salonganes, kollegaenes

Umbrella Tag	Complete Tag	Description	Examples
	Nn-Def-Sg	Definite singular noun	flyet, klokka
	Nn-Def-Sg-Gen	Genitive definite singular noun	selskapets, stovas
	Nn-Indef-Pl	Indefinite plural noun	plassar, mapper
	Nn-Indef-Pl-Gen	Genitive indefinite plural noun	tiders
	Nn-Indef-Sg	Indefinite singular noun	stol, lekam
	Nn-Indef-Sg-Gen	Genitive indefinite singular noun	fridoms
	Nn-Indef-SP	Indefinite singular or plural noun	lys, skritt
	Nn-Letter	Lowercase and uppercase letters	b, N
	Nn-Net	URL and e-mail address	www.inxight.com, info@inxight.com
Num	Num	Cardinal numeric expression or plural cardinal number (spelled out)	-294, 4,6%, xii, 1.100 to, tre
	Num-Def-Sg	Definite singular cardinal number "one" (spelled out)	eine
	Num-Indef-Sg	Indefinite singular cardinal number (spelled out)	eitt
Ord	Ord	Ordinal number (in digits or spelled out)	7., første
Part	Part-Inf	Infinitival particle	å kalla
Prep	Prep	Preposition	med, ut, opp
	Prep-av	Preposition av	av
	Prep-for	Preposition for	for
	Prep-fra	Preposition frå	frå sin opposisjon
	Prep-i	Preposition i	i

Umbrella Tag	Complete Tag	Description	Examples
	Prep-paa	Preposition på	på alle
	Prep-ved	Preposition ved	ved
Pron	Pron-Acc	Accusative pronoun	henne
	Pron-Nom	Nominative pronoun	han, ho
	Pron-Poss-Pl	Possessive pronoun with plural agreement	sine
	Pron-Poss-Sg	Possessive pronoun with singular agreement	sin
Prop	Prop	Proper name	Johan
	Prop-Gen	Genitive proper name	Espens
Punct	Punct	Miscellaneous punctuation	- [>
	Punct-Comma	Comma	,
	Punct-Quote	Quotation marks	"' '<< >> "'
	Punct-Sent	Sentence boundary punctuation	... ? ; !
V	V-Impv	Imperative verb	speil, kann
	V-Inf	Infinitive verb	gjera
	V-Inf-SForm	S-Form of infinitive verb	belastast, synast
	V-PaPart	Past participle verb	peika
	V-PaPart-SForm	S-Form of past participle verb	trivest
	V-Past	Past tense verb	sa
	V-Past-SForm	S-Form of past tense verb	møttest, mistreivst, syntest
	V-Pres	Present tense verb	gir, oppfattar
	V-Pres-SForm	S-Form of present tense verb	finst
	V-PrPart	Present participle verb	seg nemnande

4.5.5.23 Polish Part-of-Speech Tagging

Shows the tag set available for defining Polish custom entities. The tag names are accompanied by a brief description and one or more examples.

Umbrella tag	Complete Tag	Description	Examples
Adj	Adj-Sg-Acc	Singular adjective in accusative case	antydooczną, prosypialnianą, pronurkowego
	Adj-Sg-Dat	Singular adjective in dative case	zadeszczonemu, niesprzeczonej, antysylwestrowej
	Adj-Sg-Gen	Singular adjective in genitive case	zadymkowej, chudawej, antywpatowej
	Adj-Sg-Ins	Singular adjective in instrumental case	prosamorządnym, nieargusowym, niekoszykarskim
	Adj-Sg-Loc	Singular adjective in locative case	prozgnębionym, antyzniewieściatym, flankowej
	Adj-Sg-Nom	Singular adjective in nominative case	bratniem, pronajtchórzliwszy, progórkowata
	Adj-Sg-Voc	Singular adjective in vocative case	niekolbuszowskie, domyślna, antyłańcuchowa
	Adj-Pl-Acc	Plural adjective in accusative case	propośredniych, proarchiwalnych, antybrukwiane
	Adj-Pl-Dat	Plural adjective in dative case	nietrefnym, prosośnianym, antydwudziestojednodniowym
	Adj-Pl-Gen	Plural adjective in genitive case	nienaprzemiennych, antyoraczowych, przebeżowych
	Adj-Pl-Ins	Plural adjective in instrumental case	proortotonicznymi, antykaszkowymi, pragmatycznymi
	Adj-Pl-Loc	Plural adjective in locative case	antyprestwiańskich, niegorszych, śniadoliych
	Adj-Pl-Nom	Plural adjective in nominative case	lapidarniśmy, telekinetyczni, przewrotniejsi
	Adj-Pl-Voc	Plural adjective in vocative case	antypótpustynne, antypotasowi, strukturalizacyjne

Umbrella tag	Complete Tag	Description	Examples
	Adj-Inv	Adjective case and number invariant	antynawodno-, byrońsko-, prokońsko-
Adv	Adv	Adverb	wczesnojesiennie, antynieudolnie, nieusłużnie
Conj	Conj	Conjunction	i, kiedy, a, żeby
Interj	Interj	Interjection	pa, brrum, człap-człap
Nn	Nn-Sg-Acc	Singular noun in accusative case	zagraniczną, strawestowanie, napotykanie
	Nn-Sg-Dat	Singular noun in dative case	metaldehydowi, muriemu, zaniknięciu
	Nn-Sg-Gen	Singular noun in genitive case	Tanatosy, poprzeszania, podkolorowania
	Nn-Sg-Ins	Singular noun in instrumental case	wtryskiem, telemetrią, erio-metrem
	Nn-Sg-Loc	Singular noun in locative case	skodzie, sprobowaniu, przeczuciu
	Nn-Sg-Nom	Singular noun in nominative case	liczba, dziób, moralizacja
	Nn-Sg-Voc	Singular noun in vocative case	nastrojowcu, ciepolecznictwo, krukowanie
	Nn-Pl-Acc	Plural noun in accusative case	liniowania, paraliżowania, wyauskultowania
	Nn-Pl-Dat	Plural noun in dative case	szarzyznom, gonieniom, niewytwnościom
	Nn-Pl-Gen	Plural noun in genitive case	solarygrafów, zbetań, skr
	Nn-Pl-Ins	Plural noun in instrumental case	pepsynami, wykwiwami, Klimkami
	Nn-Pl-Loc	Plural noun in locative case	żurawinach, odpasowaniach, zdzierkach
	Nn-Pl-Nom	Plural noun in nominative case	krechtania, fikołki, klisze
Nn-Pl-Voc	Plural noun in vocative case	bóżnice, kantalupy, reprzy	

Umbrella tag	Complete Tag	Description	Examples
	Nn-Acc	Noun in accusative case, number invariant	ASR, CBOS, NIK
	Nn-Dat	Noun in dative case, number invariant	KRN-owi, MKS-owi, LOK-owi
	Nn-Gen	Noun in genitive case, number invariant	ZWM-u, FOZZ-u, AWF-u
	Nn-Ins	Noun in instrumental case, number invariant	PGR-em, SPEC-em, LOT- emm
	Nn-Loc	Noun in locative case, number invariant	PGR-ze, KPZR-ze, PRL-u
	Nn-Nom	Noun in nominative case, number invariant	AWF, GUC, PPR
	Nn-Sg	Singular noun, case invariant	album, parabellum, indywid- uum
	Nn-Inv	Noun case and number invariant	BGz, Kassel, NSZZ
Num	Num	Numeral	316617398,2.9, 4886, dwus- toma2
Part	Part	Particle	tuż, ponadto, nikoguteńkol
Prep	Prep	Preposition	na, ob, od, o, temu
Pron	Pron-Pers-Sg-Acc	Personal pronoun, accusative case, singular	je, mnie, cię
	Pron-Pers-Sg-Dat	Personal pronoun, dative case, singular	niemu, mnie
	Pron-Pers-Sg-Gen	Personal pronoun, genitive case, singular	jej, go, niego
	Pron-Pers-Sg-Ins	Personal pronoun, instrumental case, singular	nim
	Pron-Pers-Sg-Loc	Personal pronoun, locative case, singular	nim
	Pron-Pers-Sg-Nom	Personal pronoun, nominative case, singular	ty, ja

Umbrella tag	Complete Tag	Description	Examples
	Pron-Pers-Sg-Voc	Personal pronoun, vocative case, singular	on, ty
	Pron-Pers-Pl-Acc	Personal pronoun, accusative case, plural	je, was, ich
	Pron-Pers-Pl-Dat	Personal pronoun, dative case, plural	im, nam
	Pron-Pers-Pl-Gen	Personal pronoun, genitive case, plural	was, ich, nas
	Pron-Pers-Pl-Ins	Personal pronoun, instrumental case, plural	nimi, nami, wami
	Pron-Pers-Pl-Loc	Personal pronoun, locative case, plural	nas, was, nich
	Pron-Pers-Pl-Nom	Personal pronoun, nominative case, plural	my, one
	Pron-Pers-Pl-Voc	Personal pronoun, vocative case, plural	wy, my
	Pron-Sg-Acc	Possessive, relative, interrogative or demonstrative pronoun, accusative case, singular	ową, siakikolwiek, tamten
	Pron-Sg-Dat	Possessive, relative, interrogative or demonstrative pronoun, dative case, singular	temuż, czyjemuże, samemu
	Pron-Sg-Gen	Possessive, relative, interrogative or demonstrative pronoun, genitive case, singular	takiejże, tamtakowego, siakiejś
	Pron-Sg-Ins	Possessive, relative, interrogative or demonstrative pronoun, instrumental case, singular	onaż, niejaka, tamowęże
	Pron-Sg-Loc	Possessive, relative, interrogative or demonstrative pronoun, locative case, singular	ówż, czyjejże, tamtakowym

Umbrella tag	Complete Tag	Description	Examples
	Pron-Sg-Nom	Possessive, relative, interrogative or demonstrative pronoun, nominative case, singular	tamówż, czyje, sameż
	Pron-Sg-Voc	Possessive, relative, interrogative or demonstrative pronoun, vocative case, singular	taki, tamoweże, niektóżyż
	Pron-Voc	Possessive, relative, interrogative or demonstrative pronoun, accusative case, plural	wszystko, wszystkich, wszystkie
	Pron-Pl-Acc	Possessive, relative, interrogative or demonstrative pronoun, accusative case, plural	czyjeże, niektóreś, takichże
	Pron-Pl-Dat	Possessive, relative, interrogative or demonstrative pronoun, dative case, plural	niczymż, czyim, jakimż
	Pron-Pl-Gen	Possessive, relative, interrogative or demonstrative pronoun, genitive case, plural	tamtakowych, takichż, tamtychże
	Pron-Pl-Ins	Possessive, relative, interrogative or demonstrative pronoun, instrumental case, plural	jakimi, czyimi, tamowymż
	Pron-Pl-Loc	Possessive, relative, interrogative or demonstrative pronoun, locative case, plural	owakichże, owychż, mych
	Pron-Pl-Nom	Possessive, relative, interrogative or demonstrative pronoun, nominative case, plural	niejacy, nasi, jakieże
	Pron-Pl-Voc	Possessive, relative, interrogative or demonstrative pronoun, vocative case, plural	niejacyże, niektórekolwiek, tamte
	Pron-Acc	Possessive, relative, interrogative or demonstrative pronoun, accusative case, invariant in number	nikolwiek, nikołwiek, nikołgoż

Umbrella tag	Complete Tag	Description	Examples
	Pron-Dat	Possessive, relative, interrogative or demonstrative pronoun, dative case, invariant in number	czemukolwiek, czemukolwiek, nikomuś
	Pron-Gen	Possessive, relative, interrogative or demonstrative pronoun, genitive case, invariant in number	czegokolwiek, wszystkiego, nikogokolwiek
	Pron-Ins	Possessive, relative, interrogative or demonstrative pronoun, instrumental case, invariant in number	kim, wszystkim, wszystkimi
	Pron-Loc	Possessive, relative, interrogative or demonstrative pronoun, locative case, invariant in number	niczymś, niczym, nikim
	Pron-Nom	Possessive, relative, relative, interrogative or demonstrative pronoun, nominative case, invariant in number	niktkolwiek, ktoż, wszystko
	Pron-Voc	Possessive, relative, interrogative or demonstrative pronoun, vocative case, invariant in number	wszystko, wszystkich, wszystkie
	Pron-Refl	Reflective pronoun	sobą, siebie
Symb	Symb	Symbol	ó, ś, ”
Verb	V-Sg1	Finite verb, singular, 1st person	pooblepiałam, oddukałam, doczepiam
	V-Sg2	Finite verb, singular, 2nd person	przełamateś, przykuj, odkaszlnąteś
	V-Sg3	Finite verb, singular, 3rd person	poobrywa, przyozdabiał, zamatował
	V-P11	Finite verb, plural, 1st person	wylewałybyśmy, dosnuwamy, omierzałybyśmy
	V-P12	Finite verb, plural, 2nd person	pokraśnijcie, stypulowalibyscie, zahipotekowalibyscie

Umbrella tag	Complete Tag	Description	Examples
	V-P13	Finite verb, plural, 3rd person	zawagonują, podburzałyby, rozłasowali
	V-Inf	Verb infinitive	zatwierdzać, obśmiewać, zdrenować
	V-Impers	Impersonal verb	szkoda, brakuje, ckniiłoby
	V-Aux	Auxiliary verb (infinitive or impersonal)	móc, być, chcieć
	V-Aux-FutPart-Pl	Past tense of the auxiliary verb when it's a part of future composite tense, plural	musieli, potrafili, potrafili
	V-Aux-P11	Finite verb, plural, 1st person	potrafimy, mogłybyśmy, chcielibyśmy
	V-Aux-P12	Finite verb, plural, 2nd person	zostaniecie, musieliście, powinniście
	V-Aux-P13	Finite verb, plural, 3rd person	chcieliby, byliby, byli
	V-Aux-FutPart-Sg	Past tense of the auxiliary verb when it's a part of future composite tense, singular	potrafił, musiał, potrafił
	V-Aux-Sg1	Finite verb, singular, 1st person	byłabym, chciałam, moge
	V-Aux-Sg2	Finite verb, singular, 2nd person	byeś, musiałabyś, byłabyś
	V-Aux-Sg3	Finite verb, singular, 3rd person	mogłaby, mógł, był
	V-PaPart	Passive participle, invariable	zrąbywano, wytaplano, za-baraszkowano
	V-PaPart-Acc	Passive participle, accusative case	pierzchany, obtapiane, wite
	V-PaPart-Dat	Passive participle, dative case	synchronizowanej, przehasanemu, rozdenerwowanym
	V-PaPart-Gen	Passive participle, genitive case	rozmydlonych, skowanego, zagrożonej

Umbrella tag	Complete Tag	Description	Examples
	V-PaPart-Ins	Passive participle, instrumental case	uprawianymi, zakamienionymi, speszonymia
	V-PaPart-Loc	Passive participle, locative case	zgorzknionej, zakopcowanym, markotnionym
	V-PaPart-Nom	Passive participle, nominative case	doważony, zadarniane, leżakowana
	V-PaPart-Voc	Passive participle, vocative case	miauknięci, pozagryzani, parlowanin
	V-ActPart-Acc	Active participle, accusative case	zrzekającego, harfujące, szczerbiącą
	V-ActPart-Dat	Active participle, dative case	głaszczącej, odprzedającej, proszkującej
	V-ActPart-Gen	Active participle, genitive case	rozcholerującej, ześrodkowującej, oznajmującego
	V-ActPart-Ins	Active participle, instrumental case	rozczłonkującymi, pieszczącymi, wrębiającą
	V-ActPart-Loc	Active participle, locative case	wpuszczającej, piszczącym, ctapiących
	V-ActPart-Nom	Active participle, nominative case	hydrogenizujący, uziemiający, szalująca
	V-ActPart-Voc	Active participle, vocative case	zakrawająca, odcedzający, dębiąca
	V-PrPart	Present participle, indeclinable	ukochając, ukonkretniając, podrzemując
	V-PerfPart	Perfect participle, indeclinable	narzuciwszy, zdmuchawszy, pokropiwszy
	V-FutPart-Pl	Past tense of the verb when it's a part of future composite tense, plural	powyprzedzały, poszturchały, baletmistrzowały
	V-FutPart-Sg	Past tense of the verb when it's a part of future composite tense, singular	mumifikowało, dosadzało, zakańczało
Punct	Punct	Punctuation mark, end of sentence	:"}-]

Umbrella tag	Complete Tag	Description	Examples
	Punct-Comma	Comma	,
	Punct-Sent	Punctuation mark, end of sentence	! . ?

4.5.5.24 Portuguese Part-of-Speech Tagging

Shows the tag set available for defining Portuguese custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj	Invariant adjective	simples
	Adj-Pl	Plural adjective	cidadãos portugueses
	Adj-Sg	Singular adjective	continente europeu
Adv	Adv	Adverb	directamente
	Adv-Deg	Adverb that can modify an adjective	mais livre
Aux	Aux-be	The auxiliaries ser and estar	são, estão
	Aux-have	The auxiliaries ter and haver ('have')	tem, haverá
	Aux-Inf-be	Infinitive auxiliary estar	estar
	Aux-Inf-have	Infinitive form of ter and haver	ter, haver
	Aux-Inf-Pron-be	Infinitive 'be' auxiliary with attached clitic	sê-lo
	Aux-Inf-Pron-have	Infinitive of ter and haver with attached clitic	ter-se
	Aux-Pron-be	A 'be' auxiliary with attached clitic	estava-me
	Aux-Pron-have	ter and haver auxiliary with attached clitic	tinham-se

Umbrella Tag	Complete Tag	Description	Examples
Conj	Conj	Unclassified conjunction	nem, aquando, tal como
	Conj-Comp	Comparison conjunction	mais do que uma vez
	Conj-Coord	Coordinating conjunction	por fax ou correio
	Conj-Sub	Subordinating conjunction	para que, se, que
Det/Pron	Det/Pron-Dem-Pl	Plural demonstrative determiner or pronoun	estes, aqueles
	Det/Pron-Dem-Sg	Singular demonstrative determiner or pronoun	este, aquele
	Det/Pron-Poss-Pl	Plural possessive determiner or pronoun	vossos, seus
	Det/Pron-Poss-Sg	Singular possessive determiner or pronoun	vosso, seu
	Det/Pron-Quant-Pl	Plural quantifying determiner or pronoun	quantas vezes
	Det/Pron-Quant-Sg	Singular quantifying determiner or pronoun	quanta vez
Det	Det-Int	Interrogative determiner	demonstra a que ponto
	Det-Int-Pl	Plural interrogative determiner	quantos, quantas, quais
	Det-Int-Sg	Singular interrogative determiner	quanto, quanta, qual
	Det-Pl	Plural determiner	os maiores aplausos
	Det-Rel-Pl	Plural relative determiner	cujas
	Det-Rel-Sg	Singular relative determiner	cuja
	Det-Sg	Singular determiner	o livro
Interj	Interj	Interjection or onomatopoeia	oh, claro
Nn	Nn	Invariant noun	caos

Umbrella Tag	Complete Tag	Description	Examples
	Nn-Letter	Lowercase and uppercase letters, by themselves or followed by a period or right parenthesis	b, N
	Nn-Net	URL and e-mail address	www.inxight.cominfo@inxight.com
	Nn-Pl	Plural noun	serviços
	Nn-Sg	Singular noun	esta rede
Num	Num	Numeric expression	123
Part	Part-Neg	Negation particle	nunca
Prep	Prep	Preposition	com
	Prep-a	Preposition a	a
	Prep-Adv	Combination preposition and adverb	venho daqui
	Prep-de	Preposition de	de
	Prep-Dem-Pl	Combination preposition and plural demonstrative	desses recursos
	Prep-Dem-Sg	Combination preposition and singular demonstrative	nesta placa
	Prep-Det-Pl	Combination preposition and plural determiner	nas , longe das
	Prep-Det-Pl-a	Combination a and plural determiner	aos
	Prep-Det-Pl-de	Combination de and plural determiner	dos grandes bancos
	Prep-Det-Sg	Combination preposition and singular determiner	na construção
	Prep-Det-Sg-a	Combination a and singular determiner	ao
	Prep-Det-Sg-de	Combination de and singular determiner	da, doutro

Umbrella Tag	Complete Tag	Description	Examples
	Prep-para	Preposition para	para
	Prep-Pron	Combination preposition and pronoun	atrás dela
	Prep-Quant-Pl	Combination preposition and plural quantifier	nuns terrenos
	Prep-Quant-Sg	Combination preposition and singular quantifier	numa nuvem
	Prep-Rel	Combination preposition and relative pronoun	nesta praia aonde ...
	Prep-Rel-Pl	Combination preposition and plural relative pronoun	alunos aos quais
	Prep-Rel-Sg	Combination preposition and singular relative pronoun	área através da qual
Pron	Pron	Invariant pronoun	si
	Pron-Int-Pl	Plural interrogative pronoun	Quais são os livros de Manuel?
	Pron-Int-Sg	Singular interrogative pronoun	Qual é o livro dela?
	Pron-Pl	Plural pronoun	eles
	Pron-Rel	Invariant relative pronoun	um ortopedista que
	Pron-Rel-Pl	Plural relative pronoun	as instalações as quais
	Pron-Rel-Sg	Singular relative pronoun	o ensaio o qual
	Pron-Sg	Singular pronoun	ele
Prop	Prop	Proper noun	Lisboa, Windows
Punct	Punct	Other punctuation	: ()
	Punct-Comma	Comma	,
	Punct-Sent	Sentence punctuation	. ! ? ;
V/Adj	V/Adj-PaPart	Past participle verb or adjective	penetrado, referida

Umbrella Tag	Complete Tag	Description	Examples
V	V-Fin	Finite verb	corresponde
	V-Fin-Pron	Finite verb with attached clitic	deu-lhe
	V-Inf	Infinitive verb	reunir, conserva
	V-Inf-Pron	Infinitive verb with attached clitic	datar-se
	V-PrPart	Present participle verb	falandu
	V-PrPart-Pron	Present participle verb with attached clitic	deixandu-a

4.5.5.25 Romanian Part-of-Speech Tagging

Shows the tag set available for defining Romanian custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**.

Umbrella tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	Au, dv., fiz.
Adj	Adj	Adjective	standard, nepereche
	Adj-Indef	Adjective, indefinite	disprețuitoare, egalizatoare, behăitoare
	Adj-Pl-Indef	Adjective, plural, indefinite	galo-romane, tricotați, răsofoite
	Adj-Pl-Dir-Def	Adjective, plural, direct, definite	consternații, atenienele
	Adj-Pl-Obl-Def	Adjective, plural, oblique, definite	vertiginoaselor, neburnițateilor, interiorizațiilor
	Adj-Sg-Dir-Def	Adjective, singular, direct, definite	tâlcuiții, anahoreticul, proto-nicul
	Adj-Sg-Dir-Indef	Adjective, singular, direct, indefinite	neetichetată, dotată, neefectuată

Umbrella tag	Complete Tag	Description	Examples
	Adj-Sg-Indef	Adjective, singular, indefinite	nediversificat, neconsistat, nelefterit
	Adj-Sg-Obl-Def	Adjective, singular, oblique, definite	mercantilei, deziluzionatului, necernitului
	Adj-Sg-Obl-Indef	Adjective, singular, oblique, indefinite	monopoliste
	Adj-Sg-Voc-Def	Adjective, singular, vocative, definite	smucito, elementaro, durer-oaso
	Adj-Sg-Voc-Indef	Adjective, singular, vocative, indefinite	dragă
Adv	Adv	Adverb	devala, mâine, alocuri
Art	Art-Poss-Pl	Article, definite or possessive, plural	de-alor, ai, alor
	Art-Poss-Sg	Article, definite or possessive, singular	a
	Art-Pl-Dir	Article, non-possessive, plural, direct	cei
	Art-Pl-Obl	Article, non-possessive, plural, oblique	unor, celor
	Art-Sg-Dir	Article, non-possessive, singular, direct	o, -o, cea
	Art-Sg-Obl	Article, non-possessive, singular, oblique	unui, celei, unei
Conj	Conj	Conjunction	și, sau, dacă, deși, fiindcă
Det	Det-Pl	Pronoun or determiner, possessive or emphatic, plural	însevă
	Det-Sg	Pronoun or determiner, possessive or emphatic, singular	zostru, însuși, -mii
Interj	Interj	Interjection	tac-tac, clang, sâc
Nn	Noun	Noun, invariant	Dickens, Smith, Ampleforth
	Nn-Indef	Noun, indefinite	umezitoare

Umbrella tag	Complete Tag	Description	Examples
	Nn-Pl-Dir-Def	Noun, plural, direct, definite	echilibroarele, decalcificările, oltencele
	Nn-Pl-Indef	Noun, plural, indefinite	socoteli, necinstiri, fandosiri
	Nn-Pl-Obl-Def	Noun, plural, oblique, definite	veniturilor, supraveghetoreselor, conversiunilor
	Nn-Sg-Dir-Def	Noun, singular, direct, definite	moșmoana, esofagul, asfixia
	Nn-Sg-Dir-Indef	Noun, singular, direct, indefinite	ticăire, croială
	Nn-Sg-Indef	Noun, singular, indefinite	strop
	Nn-Sg-Obl-Def	Noun, singular, oblique, definite	salinității, listei, lopățelei
	Nn-Sg-Obl-Indef	Noun, singular, oblique, indefinite	imponderabilități
	Nn-Sg-Voc-Def	Noun, singular, vocative, definite	tunato, cumetrio, reorganizatorio
	Nn-Sg-Voc-Indef	Noun, singular, vocative, indefinite	Eugene
Num	Num	Numeral	miliardime, paisprezece-treizeci, șaispelea
Part	Part-Fut	Future particle	o fi
	Part-Infin	Infinitival particle	fără a fi
	Part-Neg	Negative particle	nu
	Part-Subj	Subjunctive particle	să
Prep	Prep	Preposition	după
Pron	Pron-Dem-Pl-Dir	Pronoun or determiner, demonstrative, plural, direct	aceiași, ceilalți, acele
	Pron-Dem-Pl-Obl	Pronoun or determiner, demonstrative, plural, oblique	acelorași, acestor
	Pron-Dem-Sg-Dir	Pronoun or determiner, demonstrative, singular, direct	același, aceea

Umbrella tag	Complete Tag	Description	Examples
	Pron-Dem-Sg-Obl	pronoun or determiner, demonstrative, singular, oblique	ăluia, cesteilalte, acestuia
	Pron-Pers-Pl-Acc	Personal pronoun, plural, accusative	va invitam
	Pron-Pers-Pl-Dat	Personal pronoun, plural, dative	ni l-a acordat
	Pron-Pers-Pl-Dir	Personal pronoun, plural, direct	ei, însele, dumnealor
	Pron-Pers-Pl-Obl	Personal pronoun, plural, oblique	dânșilor, dânselor, lor
	Pron-Pers-Sg-Acc	Personal pronoun, singular, accusative	m-, mine
	Pron-Pers-Sg-Dat	Personal pronoun, singular, dative	ție, i
	Pron-Pers-Sg-Dir	Personal pronoun, singular, direct	dumneaei, dânsa, iel
	Pron-Pers-Sg-Nom	Personal pronoun, singular, nominative	matale, eu, tu
	Pron-Pers-Sg-Obl	Personal pronoun, singular, oblique	dumitale, dumisale, lui
	Pron-Quant	Quantifier pronoun or determiner (Indefinite or negative)	câtva, vreunii
	Pron-Refl-Acc	Reflexive pronoun, accusative	se
	Pron-Refl-Dat	Reflexive pronoun, dative	sie
	Pron-Rel-Dir	Pronoun or determiner, relative, direct	câți
	Pron-Rel-Obl	Pronoun or determiner, relative, oblique	cărei
Punct	Punct-Sent	Punctuation mark, end of sentence	.!?
	Punct-Comma	Comma	,

Umbrella tag	Complete Tag	Description	Examples
	Punct	Punctuation mark	: " } -]
V	V-1	Verb, main, 1st person	săpunim, chicotesc, încrucișarăm
	V-2	Verb, main, 2nd person	vezi
	V-3	Verb, main, 3rd person	ghiorțâie, îndatoraseră, hărățea
	V-Aux-1	Verb, auxiliary, 1st person	eram
	V-Aux-3	Verb, auxiliary, 3rd person	ar
	V-Aux-P11	Verb, auxiliary, 1st person, plural	vom
	V-Aux-P12	Verb, auxiliary, 2nd person, plural	ati facut
	V-Aux-P13	Verb, auxiliary, 3rd person, plural	erau
	V-Aux-Sg1	Verb, auxiliary, 1st person, singular	nu as indrazni
	V-Aux-Sg2	Verb, auxiliary, 2nd person, singular	oi
	V-Aux-Sg3	Verb, auxiliary, 3rd person, singular	ieste
	V-Ger	Verb, gerund	înființând, gelificându, nemaiizolându
	V-Inf	Verb, infinitive	noi vrea asculta
V-Part	Verb, participle	asemuitu	

4.5.5.26 Russian Part-of-Speech Tagging

Shows the tag set available for defining Russian custom entities. The tag names are accompanied by a brief description and one or more examples. The tag set makes no distinction for number or gender.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj-Nom	Adjective in nominative case	красивый, красивая, красивое, красивые
	Adj-Acc	Adjective in accusative case	красивого, красивую, красивое, красивые
	Adj-Gen	Adjective in genitive case	красивого, красивой, красивых
	Adj-Obl	Adjective in oblique case (dative, instrumental, and so on.)	красивым, красивой, красивому, красивыми
	Adj-Comp	Comparative adjective	краше
	Adj-Brf	Adjective in brief form	красив, красива, красивы
	Adj-Inv	Abbreviated adjective	т.н.
Adv	Adv	Adverb	быстро
	Adv-Comp	Comparative adverb	лучше
Aux	Aux	Auxiliary verb	быть
Conj	Conj	Conjunction	и, но, чтобы
Det	Det-Nom	Pronominal adjective in nominative case	этот
	Det-Acc	Pronominal adjective in accusative case	эту
	Det-Gen	Pronominal adjective in genitive case	нашей
	Det-Obl	Pronominal adjective in oblique case	этому
	Det-Inv	Abbreviated pronominal adjective	др.
Dig	Dig	Number (in digits)	1999, 100M6

Umbrella Tag	Complete Tag	Description	Examples
Interj	Interj	Interjection	ага, ах, ба
Nn	Nn-Nom	Noun in nominative case	сестра, сестры
	Nn-Acc	Noun in accusative case	сестру, сестер
	Nn-Gen	Noun in genitive case	сестер
	Nn-Obl	Noun in oblique case	сестрой, сестрами
	Nn-Inv	Abbreviated noun	пр., о., г.
Num	Num	Number	три, восемь
Ord	Ord	Ordinal number (in digits)	7, 3.
Part	Part	Particle	аж, же
	Part-Int	Introduction particle	авось
	Part-Sent	Sentence particle	аминь
	Part-Mood	Mood marker particle	бы, ли
Prep	Prep-Nom	Preposition governing nominative case	плюс, минус
	Prep-Acc	Preposition governing accusative case	за
	Prep-Gen	Preposition governing genitive case	без, накануне
	Prep-Obl	Preposition governing oblique case	благодаря, к
Pron	Pron-IntRel-Nom	Relative pronoun in nominative case	кто
	Pron-IntRel-Acc	Relative pronoun in accusative case	кого
	Pron-IntRel-Gen	Relative pronoun in genitive case	чего
	Pron-IntRel-Obl	Relative pronoun in oblique case	кому

Umbrella Tag	Complete Tag	Description	Examples
	Pron-Pers-Nom	Personal pronoun in nominative case	я, ты
	Pron-Pers-Acc	Personal pronoun in accusative case	меня, тебя
	Pron-Pers-Gen	Personal pronoun in genitive case	меня, тебя
	Pron-Pers-Obl	Personal pronoun in oblique case	мною, тобой
	Pron-Adv	Pronominal adverb	откуда, кое-как
	Pron-Nom	Pronoun in nominative case	все, ничто
	Pron-Acc	Pronoun in accusative case	все
	Pron-Gen	Pronoun in genitive case	всего, ничего
	Pron-Obl	Pronoun in oblique case	всеми, ничем
Prop	Prop-Nom	Proper name in nominative case	Москва, Мальцев
	Prop-Acc	Proper name in accusative case	Москву
	Prop-Gen	Proper name in genitive case	Москвы
	Prop-Obl	Proper name in oblique case	Москве, Мальцеве
Punct	Punct-Comma	Comma	,
	Punct-Sent	Punctuation symbol at the end of a sentence	. ? !
	Punct-Symbol	Any separator in a sentence	% / \$
Verb	Verb-Fin	Finite verb	делай, делает, делал
	Verb-Ger	Adverbial participle (gerund)	делав, делавши, делая
	Verb-Inf	Infinitive verb	делать
	Verb-Acc	Participle in accusative case	делавшего, делавшую
	Verb-Gen	Participle in genitive case	делавшего, делавшей

Umbrella Tag	Complete Tag	Description	Examples
	Verb-Nom	Participle in nominative case	делавшиј, делавшее, делавшая
	Verb-Obl	Participle in oblique case	делавшим, делавшей
	Verb-Brf	Participle in brief form	делано, делана
	Verb-Inv	Abbreviated verb	исп.

4.5.5.27 Serbian (Cyrillic) Part-of-Speech Tagging

Shows the tag set available for defining Serbian custom entities (using the Cyrillic character set). The tag names are accompanied by a brief description and one or more examples. The tags are the same for both character sets.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	СНС, САД
Adj	Adj	Adjective	оштре, суров, затворен
	Adj-Comp	Comparative adjective	безбедније, ведрији
	Adj-Sup	Superlative adjective	најбољим, најмањи, највиших
Adv	Adv	Adverb	хитро, недовољно, заједно
	Adv-Comp	Comparative adverb	боље, нејасније
Conj	Conj	Conjunction	да, и, како
Enum	Enum	Enumeration	итд.
Interj	Interj	Interjection	ли, не
Nn	Nn-Pl-Nom	Plural nominative noun	очи, мисли, непријатељи
	Nn-Pl-Acc	Plural accusative noun	људе, леђа
	Nn-Pl-Gen	Plural genitive noun	очију, цифара
	Nn-Pl-Case	Plural, including vocative, locative and instrumental noun	часовницима, брковима

Umbrella Tag	Complete Tag	Description	Examples
	Nn-Sg-Nom	Singular nominative noun	ветар, плакат, струја
	Nn-Sg-Acc	Singular accusative noun	вену, лифт, плакату
	Nn-Sg-Gen	Singular, genitive noun	зида, посматрача, угла
	Nn-Sg-Case	Singular, including vocative, locative and instrumental noun	стану, природи, улици
Num	Num	Digit	14, 1984
	Num-Nom	Nominative number expression	десеторо
	Num-Acc	Accusative number expression	двоје
	Num-Case	Number expression other than nominative and accusative	троје
	Num-Card	Cardinal number	тринаест, три, једном
	Num-Ord	Ordinal number	осамнаестог, деветог, деветнаестог
Prep	Prep	Preposition	на, у, с, за
Pron	Pron	Pronoun	то, оних, све
	Pron-Pl	Plural pronoun	које, чије, никакве
	Pron-Sg	Singular pronoun	којој, чему, нечег
	Pron-Ref	Single reflexive pronoun	се
	Pron-Pers-Sg	Singular personal pronoun	њим, он, њој, те
	Pron-Pers-Pl	Plural personal pronoun	оне, вас, им
	Pron-Poss-Sg	Singular possessive pronoun	његова, мој, твој
	Pron-Poss-Pl	Plural possessive pronoun	њене, њихове, наших
Prop	Prop	Proper name	Загреб, Лондон, Евроазија
Punct	Punct	Other punctuation	...

Umbrella Tag	Complete Tag	Description	Examples
	Punct-Comma	Comma	,
	Punct-Open	Opening punctuation	(
	Punct-Close	Closing punctuation)
Verb	V-Inf	Infinitive verb	разабрати, искључити, видети
	V-Fin-Sg	Singular finite verb	избегне, посматра, обруши
	V-Fin-Pl	Plural finite verb	прате, приказују, зачуше
	V-Part	Participle	одморивши, схватајући, видевши
	V-Part-Sg	Singular participle	избијало, спречио, представљао
	V-Part-Pl	Plural participle	забијене, удешене, могле
	V-Aux-Clit	Auxiliary verb	је, би, била

4.5.5.28 Serbian (Latin) Part-of-Speech Tagging

Shows the tag set available for defining Serbian custom entities (using the Latin character set). The tag names are accompanied by a brief description and one or more examples. The tags are the same for both character sets.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	napr
Adj	Adj	Adjective	mnogima, srpskog
	Adj-Comp	Comparative adjective	dublji, manje
	Adj-Sup	Superlative adjective	najnovija, najgore
Adv	Adv	Adverb	bar, tako
	Adv-Comp	Comparative adverb	bolje, smelije
Conj	Conj	Conjunction	da, zato

Umbrella Tag	Complete Tag	Description	Examples
Enum	Enum	Enumeration	etc.
Interj	Interj	Interjection	ne, li
Nn	Nn-Pl-Nom	Plural nominative noun	srbi, ljudi
	Nn-Pl-Acc	Plural accusative noun	gorštačkim, razloge
	Nn-Pl-Gen	Plural genitive noun	svetinja, vekova
	Nn-Pl-Case	Plural, including vocative, locative and instrumental noun	funkcionerima, uslovima
	Nn-Sg-Nom	Singular nominative noun	istina, pravda
	Nn-Sg-Acc	Singular accusative noun	put, narod
	Nn-Sg-Gen	Singular, genitive noun	godine, poverenja
	Nn-Sg-Case	Singular, including vocative, locative and instrumental noun	ratu, knjizi
Num	Num	Digit	123
	Num-Nom	Nominative number expression	desetoro
	Num-Acc	Accusative number expression	dvoje
	Num-Case	Number expression other than nominative and accusative	troje
	Num-Card	Cardinal number	jedan, devet
	Num-Ord	Ordinal number	prvu, osmo
Prep	Prep	Preposition	za, od
Pron	Pron	Pronoun	svog, te
	Pron-Pl	Plural pronoun	koje
	Pron-Sg	Singular pronoun	šta
	Pron-Ref	Reflexive pronoun	se

Umbrella Tag	Complete Tag	Description	Examples
	Pron-Pers-Sg	Singular personal pronoun	mi
	Pron-Pers-Pl	Plural personal pronoun	ih
	Pron-Poss-Sg	Singular possessive pronoun	našoj
	Pron-Poss-Pl	Plural possessive pronoun	njegovih
Prop	Prop	Proper name	Zagreb
Punct	Punct	Other punctuation	...
	Punct-Sent	Sentence ending punctuation	! ? .
	Punct-Comma	Comma	,
	Punct-Open	Opening punctuation	(
	Punct-Close	Closing punctuation)
Verb	V-Inf	Infinitive verb	objasniti, uništiti
	V-Fin-Sg	Singular finite verb	reci
	V-Fin-Pl	Plural finite verb	smatraju, istaknemo
	V-Part	Participle	izvadiвши
	V-Part-Sg	Singular participle	napao, dozvolio
	V-Part-Pl	Plural participle	iskopali, proganjali
	V-Aux-Clit	Auxiliary verb	nisu, bi

4.5.5.29 Slovak Part-of-Speech Tagging

Shows the tag set available for defining Slovak custom entities. The tag names are accompanied by a brief description and one or more examples. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	dopr., hl
Adj	Adj	Adjective	úškrnových, úšustov

Umbrella Tag	Complete Tag	Description	Examples
	Adj-Comp	Comparative adjective	účtovovanší, účtovovanším
	Adj-Sup	Superlative adjective	najúbohším, najúbohších
Adv	Adv	Adverb	najavo, sami
	Adv-Comp	Comparative adverb	účastnšie, účastnejšie
	Adv-Conj	Either adverb or conjunction	ako, kde, tak
	Adv-Part	Either adverb or particle	celkom, ešte, práve
	Adv-Sup	Superlative adverb	najúbohšie, najúlisnšie
Conj	Conj	Conjunction	alebo, keby, pritom
	Conj-Part	Either conjunction or particle	a, aj, ale
Interj	Interj	Interjection	zbohom, výborne
Nn	Nn	Invariant noun	zombi, šapitó
	Nn-Pl-Gen	Plural, genitive noun	účtov, účtovaní
	Nn-Pl-Case	Plural, nominative, accusative, dative, locative and instrumental noun	účtami, účtovaniami
	Nn-Sg-Gen	Singular, genitive noun	účtu, účtovania
	Nn-Sg-Case	Singular, nominative, accusative, dative, locative and instrumental noun	účtovi, účtovaním
	Nn-Net	URL, e-mail address	www.inxight.com , info@inxight.com
Num	Num	Number expression other than cardinal or ordinal, ascii numbers	1, 12%
	Num-Card	Cardinal number	osemsto, štyritisíc osemsto
	Num-Ord	Ordinal number	dvetisíc, dvetisícsto
Part	Part	Particle	nie, by
Pref	Pref	Prefix (stand alone prefix)	vodo, ne

Umbrella Tag	Complete Tag	Description	Examples
Prep	Prep	Preposition	v, zo
Pron	Pron-Dem-Pl	Plural demonstrative pronoun	všetitakí, všetitakým
	Pron-Dem-Sg	Singular demonstrative pronoun	taký, všetitakom
	Pron	Indefinite pronoun	čosi
	Pron-Pl	Plural pronoun	dačíchsi, čiesi
	Pron-Sg	Singular pronoun	kdečiasí, všeličiasí
	Pron-Interrog	Interrogative pronoun	kto, všetikoho
	Pron-Refl	Reflexive pronoun	sám, svoj
	Pron-Pers-Sg	Singular personal pronoun	ona, on
	Pron-Pers-Pl	Plural personal pronoun	oni, ony
	Pron-Poss	Possessive pronoun	váš, ich
Prop	Prop	Proper noun	Swisscom, Swisscomami
Punct	Punct	Other punctuation	... -
	Punct-Sent	Sentence ending punctuation	! ? .
	Punct-Comma	Comma	,
	Punct-Open	Opening punctuation	(
	Punct-Close	Closing punctuation)
	Punct-Quote	Quote	"
V	V-Inf	Infinitive verb	účtovať, účtovať
	V-Past-Pl	Plural, past tense verb	účtovali, účinkovali
	V-Past-Sg	Singular, past tense verb	účtoval, účtovala
	V-Pres-Pl	Plural, present tense verb	účtovujú, účtujeme
	V-Pres-Sg	Singular, present tense verb	účtovujem, účtovuj
	V-Fut-Pl	Plural, future tense verb	budú, budete

Umbrella Tag	Complete Tag	Description	Examples
	V-Fut-Sg	Singular, future tense verb	bude, budeš
	V-Aux	Auxiliary verb	vie, vieš

4.5.5.30 Slovenian Part-of-Speech Tagging

Shows the tag set available for defining Slovenian custom entities. The tag names are accompanied by a brief description and one or more examples. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj	Adjective	miren, mirna, mlad, mladi
	Adj-Comp	Comparative adjective	lepši, lepše, bolj divji, manj divji
	Adj-Sup	Superlative adjective	najlepši, najlepše, najbolj divji, najmanj divji
Adv	Adv	Adverb	lepo, naglo, nagloma
	Adv-Comp	Comparative adverb	lepše, bolj vroče, manj razločno
	Adv-Sup	Superlative adverb	najlepše, najbolj vroče, najmanj razločno
Conj	Conj	Conjunction	in, pa, medtem ko
	Conj-Part	Conjunction or participle	samo, ne
Interj	Interj	Interjection	pfuj, ehej
Nn	Nn	Invariant noun, including abbreviations, acronyms and so on	ZDA, št.
	Nn-Du-Gen	Dual, genitive noun	čvrstosti
	Nn-Du-Case	Dual, nominative, accusative, dative, locative and instrumental noun	čvrstostih, čvrstostima
	Nn-Pl-Gen	Plural, genitive noun	čvrstosti

Umbrella Tag	Complete Tag	Description	Examples
	Nn-Pl-Case	Plural, nominative, accusative, dative, locative and instrumental noun	čvrstostih, čvrstostmi
	Nn-Sg-Gen	Singular, genitive noun	čvrstosti
	Nn-Sg-Case	Singular, nominative, accusative, dative, locative and instrumental noun	čvrstosti, čvrstostjo
	Nn-Net	URL, e-mail address	www.inxight.com, info@inxight.com
Num	Num	Invariant number expression	1, 12%
	Num-Card-Gen	Cardinal number, genitive	stotih
	Num-Card-Case	Cardinal number, nominative, accusative, dative, locative and instrumental	sto, stotim
	Num-Ord-Gen	Ordinal number, genitive	stotih
	Num-Ord-Case	Ordinal number, nominative, accusative, dative, locative and instrumental	sto, stotim
	Num-Gen	Number expression other than cardinal or ordinal numbers, genitive	čtvorke, četvork
	Num-Case	Number expression other than cardinal or ordinal, nominative, accusative, dative, locative and instrumental	čtvorka, četvorki
Part	Part	Particle	že, žal
Prep	Prep	Preposition	pod, po
	Prep-Cmpd	Preposition with clitic	podnje, podnjo
Pron	Pron-Dem-Du	Dual demonstrative pronoun	toliki, tolíkima
	Pron-Dem-Pl	Plural demonstrative pronoun	tolike, tolíkimi

Umbrella Tag	Complete Tag	Description	Examples
	Pron-Dem-Sg	Singular demonstrative pronoun	to, toliko
	Pron-Ref	Reflexive pronoun, invariant in number	sebe, seboj, sebi
	Pron-Ref-Sg	Singular reflexive pronoun	svoj, svojim, svoji
	Pron-Ref-Du	Dual reflexive pronoun	svoji, svojih, svojima
	Pron-Ref-Pl	Plural reflexive pronoun	svoji, svoje, svoja
	Pron-Pers-Sg	Singular personal pronoun	jaz, ti, on, ona, ono
	Pron-Pers-Du	Dual personal pronoun	midva, vidva, onadva
	Pron-Pers-Pl	Personal pronoun, plural	mi, me, vi, ve, oni
	Pron-Poss-Sg	Possessive pronoun, singular	moj, tvoj, njen, njegov
	Pron-Poss-Du	Possessive pronoun, dual	najin, vajin, njun
	Pron-Poss-Pl	Possessive pronoun, plural	naš, vaš, njihov
	Pron-Interrog	Interrogative pronoun	kdo, kaj, kateri
	Pron-Rel	Relative pronoun	kdor, kar, kateri, ki
	Pron-Pl	Plural pronoun	vsem, vse, vse, vsa
	Pron-Du	Dual pronoun	vsi, vse
	Pron-Sg	Singular pronoun	vso, vsm, vse
	Pron	Other pronouns, indefinite, evaluative and so on	isti, drug
Prop	Prop	Proper noun	Sava, Ljubljana Prop
Punct	Punct	Other punctuation	... -
	Punct-Sent	Sentence-ending punctuation	. ! ?
	Punct-Comma	Comma	,
	Punct-Open	Opening punctuation	(
	Punct-Close	Closing punctuation)

Umbrella Tag	Complete Tag	Description	Examples
	Punct-Quote	Quote	"
V	V-Aux	Auxiliary verb	biti, bi
	V-Sup	Supine verb	prodat, spat
	V-Inf	Infinitive verb	prodati, spati
	V-PPast-Du	Dual, past tense verb	čvrčali, čvrčala
	V-PPast-Pl	Plural, past tense verb	čvrčala, čvrčale
	V-PPast-Sg	Singular, past tense verb	čvrčalo, čvrčal
	V-Pres-Du	Dual, present tense verb	jva, jta
	V-Pres-Pl	Plural, present tense verb	jte, jmo
	V-Pres-Sg	Singular, present tense verb	je, ješ

4.5.5.31 Spanish Part-of-Speech Tagging

Shows the tag set available for defining Spanish custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj	Invariant adjective	beige, mini
	Adj-Ord-Pl	Plural, spelled out ordinal number	primeros
	Adj-Ord-Sg	Singular, spelled out ordinal number	primer, primera, sexta
	Adj-Pl	Plural adjective	bonitos, nacionales
Adv	Adv	Adverb	siempre, directamente
	Adv-Deg	Adverb that can modify an adjective	mu y importante
	Adv-Int	Interrogative adverb	cuándo

Umbrella Tag	Complete Tag	Description	Examples
	Adv-Rel	Adverbial relativizer	donde
Aux	Aux-be	The auxiliaries ser and estar ('be')	es, fui, estaba
	Aux-have	The auxiliary haber ('have')	han, hubo, hay
	Aux-Inf-be	Infinitive 'be' auxiliary	estar
	Aux-Inf-have	Infinitive form of haber	haber
	Aux-Inf-Pron-be	Infinitive be auxiliary with attached clitic	serme, estarlo
	Aux-Inf-Pron-have	Infinitive of haber with attached clitic	haberle, haberseme
Conj	Conj	Conjunction	si, porque
	Conj-como	The word como	como
	Conj-Coord	Coordinating conjunction	y, o
	Conj-que	The word que	que
Det/Pron	Det/Pron-Quant-Pl	Plural quantifying determiner or pronoun	unas casas
	Det/Pron-Quant-Sg	Singular quantifying determiner or pronoun	poca
Det	Det-Dem-Pl	Plural demonstrative determiner	estas, esos
	Det-Dem-Sg	Singular demonstrative determiner	esta
	Det-Pl	Plural determiner	tus
	Det-Pre-Pl	Plural pre-determiner	todas las, todos los
	Det-Pre-Sg	Singular pre-determiner	todo el, toda la
	Det-Rel	Relative determiner	cual, cuyo
	Det-Sg	Singular determiner	mi
Interj	Interj	Interjection or onomatopoeia	ah

Umbrella Tag	Complete Tag	Description	Examples
Nn	Nn	Noun, invariant for number	fénix
	Nn-Letter	Lowercase letter with or without a period or an uppercase without a period	a, h., M
	Nn-Net	URL or e-mail address	www.inxight.com, info@inxight.com
	Nn-Pl	Plural noun	gatos
	Nn-Sg	Singular noun	gato
Num	Num	Numeric expression, cardinal number	123, XIX, once, cuatrocientos
	Num-Ord	Ordinal number	1o., 2ª
Part	Part-Neg	The negation particle no	no
Prep	Prep	Preposition	en, con, por
	Prep-a	Preposition a	a casa
	Prep-de	Preposition de	la casa de María
	Prep-Det	Combination of preposition and determiner	fuera del , antes del
	Prep-Det-a	Combination a and determiner	al
	Prep-Det-de	Combination de and determiner	del
	Prep-para	Preposition para	para la casa
Pron	Pron	Pronoun	yo
	Pron-Clitic	Clitic pronoun (acc. or dat.)	le, me, os, nos
	Pron-Dem	Demonstrative pronoun	ésta, aquél
	Pron-Int	Interrogative pronoun	cuánto, cuál, quién
	Pron-Poss	Possessive pronoun	el mío , las vuestras
	Pron-Rel	Relative pronoun	lo cual, quien

Umbrella Tag	Complete Tag	Description	Examples
	Pron-se	The reflexive pronoun	se
Prop	Prop	Proper noun or alpha-numeric-punctuation combinations	Pablo , U-587, Win2000
Punct	Punct	Other punctuation	' " / & { ; :
	Punct-Comma	Comma	,
	Punct-Sent	Sentence punctuation	. ? !
V/Adj	V/Adj-PaPart-Pl	Plural past participle verb or adjective	hechas
	V/Adj-PaPart-Sg	Singular past participle verb or adjective	fundada
V	V-Fin	Finite verb	tiene, pueda, dicte
	V-Impv	Imperative verb	dejad
	V-Impv-Pron	Imperative verb with attached clitic	déjame, sígueme
	V-Inf	Infinitive verb	evitar, tener, conducir
	V-Inf-Pron	Infinitive verb with attached clitic	hacerse, suprimirlas
	V-PrPart	Present participle verb	siendo, tocando
	V-PrPart-Pron	Present participle verb with attached clitic	haciéndoles, tomándolas

4.5.5.32 Swedish Part-of-Speech Tagging

Shows the tag set available for defining Swedish custom entities. The tag names are accompanied by a brief description and one or more examples. If the example consists of more than one word, the word exemplifying the current tag is **emphasized**. The tag set makes no distinction for gender.

Umbrella Tag	Complete Tag	Description	Examples
Abbr	Abbr	Abbreviation	S.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj-Comp	Comparative adjective	äldre
	Adj-Def-Pl	Definite and/or plural adjective	svenska
	Adj-Indef-Sg	Indefinite singular adjective	grov
	Adj-Sup	Superlative adjective	viktigast
	Adj-Sup-Def	Definite superlative adjective	fullaste
Adv	Adv	Adverb	redan
Cmpd	Cmpd-Part	Compound part	plats- och släktnamnen
Conj	Conj	Conjunction	att
	Conj-Coord	Coordinating conjunction	och, eller
Det/Pron	Det/Pron-Pl	Plural determiner or pronoun	dessa
	Det/Pron-Pl-Gen	Plural genitive determiner or pronoun	andras
	Det/Pron-Sg	Singular determiner or pronoun	något, denna
	Det/Pron-Sg-Gen	Singular genitive determiner or pronoun	dennes
Det	Det-Def-Pl	Definite plural determiner	de partierna
	Det-Def-Sg	Definite singular determiner	den ekonom
	Det-Indef-Sg	Indefinite singular determiner	en, ett
Interj	Interj	Interjection	ja
Nn	Nn-Def-Pl	Definite plural noun	verken
	Nn-Def-Pl-Gen	Genitive definite plural noun	professionernas
	Nn-Def-Sg	Definite singular noun	historikern
	Nn-Def-Sg-Gen	Genitive definite singular noun	dagens
	Nn-Indef-Pl	Indefinite plural noun	strukturer

Umbrella Tag	Complete Tag	Description	Examples
	Nn-Indef-Pl-Gen	Genitive indefinite plural noun	rikens
	Nn-Indef-Sg	Indefinite singular noun	dag
	Nn-Indef-Sg-Gen	Genitive indefinite singular noun	institutions
	Nn-Indef-SP	Indefinite singular or plural noun	universitet
	Nn-Indef-SP-Gen	Genitive indefinite singular or plural noun	års
	Nn-Letter	Lowercase and uppercase letters	b, N
	Nn-Net	URL or e-mail address	www.inxight.com info@inxight.com
Num	Num	Cardinal number, in digits or spelled-out plural number	5,4 or 300 hundra
	Num-Sg	Spelled-out number "one"	en
Ord	Ord	Ordinal number (digits or words)	tredje
Prep	Prep	Preposition	kring
	Prep-av	Preposition av	av
	Prep-foer	Preposition för	för
	Prep-om	Preposition om	om
	Prep-paa	Preposition på	på
Pron	Pron-Acc	Accusative pronoun	en
	Pron-Gen	Genitive pronoun	ens
	Pron-Nom	Nominative pronoun	man
	Pron-Pers-Acc	Accusative personal pronoun	sig
	Pron-Pers-Gen	Genitive personal pronoun	dess, deras

Umbrella Tag	Complete Tag	Description	Examples
	Pron-Pers-Nom	Nominative personal pronoun	du
	Pron-Poss-Pl	Plural possessive pronoun	mina
	Pron-Poss-Sg	Singular possessive pronoun	vår, ert
Prop	Prop	Proper name	Europa, Margareta
	Prop-Gen	Genitive proper name	Eriks
Punct	Punct	Punctuation	/ -
	Punct-Comma	Comma	,
	Punct-Paren	Bracketing punctuation	() []
	Punct-Quote	Quotation punctuation	'' « »
	Punct-Sent	Sentence terminating punctuation	. ? !
Quant	Quant	Quantifier	alla, många
V	V-Impv	Imperative verb	tänk
	V-Impv-SForm	Imperative verb, S-Form	minns
	V-Inf	Infinitive verb	skilja
	V-Inf-SForm	Infinitive verb, S-Form	tänkas
	V-PaPart	Past participle verb	tecknat, handlat
	V-Past	Past tense verb	slog
	V-Past-SForm	Past tense verb, S-Form	kysstes
	V-Pres	Present tense verb	varnar
	V-Pres-SForm	Present tense verb, S-Form	sägs
	V-PrPart	Present participle verb	mötande

4.5.5.33 Thai Part-of-Speech Tagging

Shows the tag set available for defining Thai custom entities. The tag names are accompanied by a brief description and one or more examples.

Umbrella tag	Complete tag	Description	Examples
Adj	Adj	Adjective	เป็นธรรมชาติ , อีสระ , แข็ง
Adv	Adv	Adverb	กัน , มาก , คล่อง
Aux	Aux	Auxiliary	จะ , ต้อง , ได้
Cl	Cl	Classifier	วัน , คน , ปี
Conj	Conj	Conjunction	ไม่ว่า , ก็ , ในขณะที่
Det	Det	Determiner	นั่น , อีก , ที่นี่
Interj	Interj	Interjection	อิจฉา , โอ๊ย , โครม
Neg	Neg	Negator	ไม่ได้ , ไม่ , เปล่า
Nn	Nn	Noun	เรื่อง , กรณี , การเมือง
	Nn-Prop	Proper noun	เดลินิวส์ , ศาลปกครอง , ชัมชุง
Num	Num	Cardinal number	สอง , 3, 10.5
Ord	Ord	Ordinal number	ที่สาม , สุดท้าย , แรก
Part	Part	Particle	หรือไม่ , หน่อย , นะ
Prep	Prep	Preposition	เพื่อ , ใน , จาก
Pron	Pron	Pronoun	เจ้า , ที่ , เรา
Punct	Punct	Punctuation	(. - .)
Quant	Quant	Quantifier	ครั้ง , บาง , หลาย
Verb	Verb	Verb	เป็น , หรือ , ปฏิรูป

4.5.5.34 Turkish Part-of-Speech Tagging

Shows the tag set available for defining Turkish custom entities. The tag names are accompanied by a brief description and one or more examples.

Umbrella Tag	Complete Tag	Description	Examples
Adj	Adj	Adjective	korkunç, görünmez, yekpare, kaybettiğim, tanımadığım, oturduğum
Adv	Adv	Adverb	artık
Conj	Conj	Conjunction	ve, fakat, yahut
Det	Det	Determiner	bir, her, o
Dup	Dup	Onomatopoeic words which only appear as duplications in the sentence	pırıl, cazır
Interj	Interj	Interjection	ah, haydi
Nn	Nn-Pl-Nom	Plural noun in nominative case	şekiller, dualar, ihlaslar
	Nn-Pl-Gen	Plural noun in genitive case	şeylerin, merhalelerinin, gir-enlerin
	Nn-Pl-Dat	Plural noun in gative case	evdekilere, uzaklara, sabah-lara
	Nn-Pl-Acc	Plural noun in gccusative/ objective case	şeyleri, yazıları, kesecikleri
	Nn-Pl-Equ	Plural noun in equative case (by (object) in passive sentences)	onlarca, vatandaşlarca
	Nn-Pl-Abl	Plural noun in ablative case	çocuklardan, inananlardan
	Nn-Pl-Ins	Plural noun in instrumental case	icatlarla, kişilerle
	Nn-Pl-Loc	Plural noun in locative case	saatlerde, havalarda
	Nn-Sg-Nom	Singular noun in nominative case	nosyon, akış, dünya

Umbrella Tag	Complete Tag	Description	Examples
	Nn-Sg-Gen	Singular noun in genitive case	akışın, perdenin, yatağın
	Nn-Sg-Dat	Singular noun in dative case	araya, görmeye, hayaline
	Nn-Sg-Acc	Singular noun in accusative/objective case	şeyi, çetrefili, Mozartı
	Nn-Sg-Equ	Singular noun in equative case (by (object) in passive sentences)	çocukça, dünyaca
	Nn-Sg-Abl	Singular noun in ablative case	karanlıktan, defadan
	Nn-Sg-Ins	Singular noun in instrumental case	hayretle, dehşetle, hızla
	Nn-Sg-Loc	Singular noun in locative case	yumakta, şekilde
Num	Num	Numeral	birinci, on, yüzlerce
Postp	Postp	Postposition	beraber, kadar, sonra
Pron	Pron-Pl-Abl	Pronoun in ablative case plural	onlardan
	Pron-Pl-Acc	Pronoun in accusative case plural	onları
	Pron-Pl-Dat	Pronoun in dative case plural	onlara
	Pron-Pl-Equ	Pronoun in equative case plural	onlarca
	Pron-Pl-Gen	Pronoun in genitive case plural	onların
	Pron-Pl-Ins	Pronoun in instrumental case plural	onlarla
	Pron-Pl-Loc	Pronoun in locative case plural	onlarda
	Pron-Pl-Nom	Pronoun in nominative case plural	onlar

Umbrella Tag	Complete Tag	Description	Examples
	Pron-Sg-Abl	Pronoun in ablative case singular	ondan
	Pron-Sg-Acc	Pronoun in accusative case singular	onu
	Pron-Sg-Dat	Pronoun in dative case singular	ona lar
	Pron-Sg-Equ	Pronoun in equative case singular	onca
	Pron-Sg-Gen	Pronoun in genitive case singular	onun
	Pron-Sg-Ins	Pronoun in instrumental case singular	onla, onunla
	Pron-Sg-Loc	Pronoun in locative case singular	onda
	Pron-Sg-Nom	Pronoun in nominative case singular	o
Punct	Punct	Punctuation	, " :
	Punct-Sent	Punctuation, end of sentence	. ! ?
Ques	Ques-P11	Question word, first person plural	muyduk
	Ques-P12	Question word, second person plural	muydunuz
	Ques-P13	Question word, third person plural	muydular
	Ques-Sg1	Question word, first person singular	muydumup mu
	Ques-Sg2	Question word, second person singular	muydun
	Ques-Sg3	Question word, third person singular	muydu
V	V-P11	Plural verb, first person	yapalım, biliyoruz

Umbrella Tag	Complete Tag	Description	Examples
	V-P11-Neg	Plural verb, first person, negated	yyemedik, alamayız
	V-P12	Plural verb, second person	olursanız, katlanırsınız, dinlediniz
	V-P12-Neg	Plural verb, second person, negated	beklemeyin, tüketmeyin
	V-P13	Plural verb, third person	sönüyorlardı, sustular, şeylerdir
	V-P13-Neg	Plural verb, third person, negated	görmüyorlardı, vermiyorlardı
	V-Sg1	Singular verb, first person	duydum, başladım, görüyordum
	V-Sg1-Neg	Singular verb, first person, negated	bilmiyorum, anlamıyordum
	V-Sg2	Singular verb, second person	affet, yoksun, bahsediyorsun
	V-Sg2-Neg	Singular verb, second person, negated	çıkarma, kullanmıyorsun
	V-Sg3	Singular verb, third person	eridi, akıyordu, örtüyordu
	V-Sg3-Neg	Singular verb, third person, negated	vermiyor, belirlememeli

5 Entity Extraction

Extracting entities from unstructured text tells us what the text is about—the people, organizations, places, and other parties described in the document.

The extraction process involves processing and analyzing text, finding entities of interest, assigning them to the appropriate type, and presenting this metadata in a standard format, including the entity's character offset in the document, the entity's normalized ISO format, and other attributes.

The extraction process can extract entities using lists of specific named entities.

Entities are often proper names, such as the names of specific and unique people, organizations, or places. Other specified entity types include currency amounts and dates, among others.

Each entity is defined as a pairing of a name and its type. For example:

- Canada/COUNTRY
- John Paul/PERSON
- General Motors Corporation/ORGANIZATION/COMMERCIAL

Entity types play a crucial role in the definition of an entity. Entity types are used to classify entities extracted from documents and entities stored in a dictionary.

The software contains an extensive set of predefined entity types. You can optionally enhance the extraction process by using dictionaries and extraction rules.

For more details about creating dictionaries and extraction rules, refer to the *SAP Data Services Text Data Processing Extraction Customization Guide*.

i Note

The more general entity type `NOUN_GROUP` is meant to extract noun phrases consisting of one or more related nouns, or modifier(s) plus noun(s), which are not identified as other more specific entities, such as name, measure, and so on.

5.1 Predefined Entity Types by Language

This table lists the predefined entity types in alphabetical order and indicates which languages support them. Although not shown in the table, `NOUN_GROUP` is extracted by all language modules.

i Note

For a list of additional public sector entities, see [Public Sector Fact Extraction \[page 346\]](#).

In this table of Entity Type names and language modules to which they apply, the Chinese examples apply to both Simplified and Traditional Chinese.

Entity Type Name	Arabic	Chinese	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ADDRESS1 [page 295]	X	X	X	X	X	X	X	X	X	X	X	X	X
ADDRESS2 [page 295]			X	X		X	X	X			X	X	X
CONTINENT [page 295]	X	X	X	X		X	X	X	X	X	X	X	X
COUNTRY [page 295]	X	X	X	X	X	X	X	X	X	X	X	X	X
CURRENCY [page 296]	X	X	X	X	X	X	X	X	X	X	X	X	X
DATE [page 296]	X	X	X	X	X	X	X	X	X	X	X	X	X
DAY [page 296]	X	X	X	X	X	X	X	X	X	X	X	X	X
FACILITY [page 296]	X									X			
FACILITY@AIRPORT [page 297]	X	X		X									
FACILITY@BUILD- GROUNDS [page 297]	X	X		X									
FACILITY@PATH [page 297]	X	X		X									
FACILITY@PLANT [page 297]	X	X		X									
FACILITY@SUBAREA [page 297]	X	X		X									
GEO_AREA [page 297]	X		X			X	X	X	X	X	X	X	X

Entity Type Name	Arabic	Chinese	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
GEO_AREA@DOMESTIC [page 297]		X		X									
GEO_AREA@INTL [page 298]		X		X									
GEO_FEATURE [page 298]	X		X		X	X	X	X	X	X	X	X	X
GEO_FEATURE@BOUNDARY [page 298]		X		X									
GEO_FEATURE@CELESTIAL [page 298]		X		X									
GEO_FEATURE@LAND [page 298]		X		X									
GEO_FEATURE@WATER [page 298]		X		X									
GEOCOORD [page 298]		X		X									
HOLIDAY [page 299]	X	X	X	X	X	X	X	X	X	X	X	X	X
LANGUAGE [page 299]		X	X	X		X	X	X	X	X	X	X	X
LOCALITY [page 299]	X	X	X	X	X	X	X	X	X	X	X	X	X
MEASURE [page 299]	X	X	X	X	X	X	X	X	X	X	X	X	X
MGRS [page 299]				X									
MONTH [page 299]	X	X	X	X	X	X	X	X	X	X	X	X	X

Entity Type Name	Arabic	Chinese	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
NAME_DESIGNATOR [page 299]			X	X		X		X	X		X		X
NIN@CA_SIN [page 300]						X							
NIN@CH_AVS [page 300]								X					
NIN@FR_INSEE [page 300]						X							
NIN@IT_CF [page 300]								X					
NIN@KR_RRN [page 301]										X			
NIN@NL_BSN [page 301]			X										
NIN@US_SSN [page 301]				X									
NOUN_GROUP [page 301]	X	X	X	X	X	X	X	X	X	X	X	X	X
ORGANIZATION@COMMERCIAL [page 301]	X	X	X	X	X	X	X	X	X	X	X	X	X
ORGANIZATION@EDUCATIONAL [page 302]		X	X	X		X	X	X	X	X	X	X	X
ORGANIZATION@ENTERTAINMENT [page 302]		X		X									

Entity Type Name	Arabic	Chinese	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ORGANIZATION@GOVERNMENT [page 302]		X		X									
ORGANIZATION@MEDIA [page 302]		X		X									
ORGANIZATION@MEDICAL-SCIENCE [page 302]		X		X									
ORGANIZATION@RELIGIOUS [page 302]		X		X									
ORGANIZATION@SPORTS [page 302]		X		X									
ORGANIZATION@OTHER [page 303]	X	X	X	X	X	X	X	X	X	X	X	X	X
PEOPLE [page 303]	X	X	X	X		X	X	X	X	X	X	X	X
PERCENT [page 303]	X	X	X	X	X	X	X	X	X	X	X	X	X
PERSON [page 303]	X	X	X	X	X	X	X	X	X	X	X	X	X
PHONE [page 303]	X	X	X	X	X	X	X	X	X	X	X	X	X
PRECURSOR@CHEMICAL [page 303]				X									
PRECURSOR@NUCLEAR [page 303]				X									
PRODUCT [page 304]		X	X	X		X	X	X	X	X	X	X	X

Entity Type Name	Arabic	Chinese	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
PROP_MISC [page 304]		X	X	X		X	X	X	X		X	X	X
REGION@MAJOR [page 304]	X	X	X	X	X	X	X	X	X	X	X	X	X
REGION@MINOR [page 305]	X		X	X		X	X	X		X		X	X
SOCIAL_MEDIA@ID_TWITTER [page 305]		X	X	X		X	X	X	X	X	X	X	X
SOCIAL_MEDIA@TOPIC_TWITTER [page 305]		X	X	X		X	X	X	X	X	X	X	X
TICKER [page 305]			X	X		X	X	X			X	X	X
TIME [page 305]	X	X	X	X	X	X	X	X	X	X	X	X	X
TIME_PERIOD [page 305]	X	X	X	X	X	X	X	X	X	X	X	X	X
TITLE [page 305]	X	X	X	X	X	X	X	X	X	X	X	X	X
URI@EMAIL [page 305]	X	X	X	X	X	X	X	X	X	X	X	X	X
URI@IP [page 306]		X	X	X		X	X	X	X	X	X	X	X
URI@URL [page 306]	X	X	X	X		X	X	X	X	X	X	X	X
VEHICLE@AIR [page 306]	X	X		X									
VEHICLE@LAND [page 306]	X	X		X									

Entity Type Name	Arabic	Chinese	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
VEHICLE@WATER [page 306]	X	X		X									
VEHICLE@LICENSE [page 306]	X			X									
VEHICLE@VIN [page 306]	X			X									
WEAPON@BIOLOGICAL [page 307]		X		X									
WEAPON@CHEMICAL [page 307]		X		X									
WEAPON@EXPLODING [page 307]		X		X									
WEAPON@NUCLEAR [page 307]		X		X									
WEAPON@PROJECTILE [page 307]		X		X									
WEAPON@SHOOTING [page 307]		X		X									
YEAR [page 307]	X	X	X	X	X	X	X	X	X	X	X	X	X

5.2 Language-specific Entity Type Examples

Examples of the entity types that can be extracted for Arabic, Chinese (Simplified & Traditional), Dutch, English, Farsi, French, German, Italian, Japanese, Korean, Portuguese, Russian, and Spanish.

In the table below, the links in the first column are to the Definitions section of the parent document, so hovering on the link displays the definition. Each predefined entity type name has the same definition, no

matter what the language. Rather than repeating the definition for each language and consuming space, we give the reader a way to view it if desired.

This table can be displayed with any of the columns hidden. Initially, only Arabic, Simplified Chinese, English, and German are shown. Use the [Show/hide columns](#) button to control what is displayed.

Entering text in the [Search](#) box causes the table to show only rows that contain that text in any column.

i Note

This search displays a row even if the text appears only in a hidden column.

Entering text in the box below a column title causes the table to display all rows that contain that text in that column.

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ADDRESS	ش صلاح الدين الاسكندرية 38 العنوان البريدي ص.ب. 68333 الرياض 11527 المملكة العربية السعودية	北京市朝阳门外大街甲12号新华保险大厦7层701室 (100022)	北京市朝阳区建国门外大街甲12号新华保险大厦7层701室 (100022)	Prinses Beatrixlaan 14 Postbus 20002 Antwoordnummer 10010 Archimedesstraat 36-38 0-58 2304 室 (200041)	Prinses Beatrixlaan 14 Postbus 20002 Antwoordnummer 10010 Archimedesstraat 36-38 0-58 2304 室 (200041)	خیابان شعرینی کوچه سوم ۸۹۹	Theseeexampl eshowcomplet eadresses asfound in ad document. The bolded content	Theseeexampl eshowcomplet eadresses asfound in ad document. The bolded content	Viale Guisone 1 Casella postale 109 Piazza Colonna, 370 Località Valtericca, 375	540-8 大阪府大阪市中央区 見2丁目2番 53号 329-1104 栃木県宇都宮市下岡本町 2145番 13号	서울시 용산구 동빙고동 28-32 삼거리 칠물실비 140-210 서울시 송파구 삼전동 33-15 서울시 노원구 상계동 1021-4 광진빌라104호 137-823	Rua Santa Maria de Itabira, 147 Bairrosion Caixa Postal 15012 SAUS Quadr a 06, Bloco F, 2º andar R. Esteves Junior, 50 - loja 13 Ed. Top	проспект Вернадского, 97 проспект Боголубова, дом 25	Theseeexampl eshowcomplet eadresses asfound in ad document. The bolded content

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
					t is th e AD DR ES S1 val ue .		t is th e AD DR ES S1 val ue .	bo ld ed co nt en tis th e AD DR ES S1 val ue .				DR ES S1 ca n pr ec ed e or fol lo w AD DR ES S2 .	t is th e AD DR ES S1 val ue :	
					1234 Ma- hana St., Hono- lulu Hawai'i 96816		27 rue Pas- teur, Sher- brooke , Qué- bec, J1K 2Y3, Can- ada	Kai- ser- straße 123, 10623 Berlin, Deutsc hland				Bo th for m at s ar e in us e in Ru ssi an - sp ea ki ng co un tri es.	Ave- nida Cristo- bal Colón 5667 Plaza de la Leal- tad, 5 28014 Madrid	
					PO Box 10101, Bos- ton, MA 02345		4, rue du 8 Mai 1945, Van- couver, BC, V6E 1R8	Stock- erauer straße 9, A8700 Leo- ben						Calle Cas- tillo Cha- pulte- pec 47 62380 Cuer- navaca , Méx- ico
					Rural Route 5, Lex- ington NE 69361		Case Post- ale 123, Suc- cur- sale Cen-	Zuger- berg- strass e 18, CH-64						i N

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish	
							tre-Ville , Montréal, PQ, Canada 31bis , Saint-Joseph nord , 13402 Marseille, France Rue du Cornet 6 , B-480 0 VERVIERS 19 , quai de la Voltaire , Paris Il habite 68bis avenue des Abb-	14 Untersee-geri Postfach 10 43 51 , D-700 49, Stuttgart, Deutschland							ot e The form at for ADDRESS1 is based on typical postal addresses patterns found in Peninsular, Mexico

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
							esses							n
							pas							an
							loin de							d
							la							So
							phar-							ut
							macie							h
														A
														m
														eri
														ca
														n
														Sp
														an
														ish
														ad
														dr
														es
														se
														s.

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
Address			2595		i N ot e Th es e ex a m pl es sh ow co m pl et e ad dr es se s as fo un d in a do cu m en t. Th e bo ld ed co nt en		i N ot e Th es e ex a m pl es sh ow co m pl et e ad dr es se s as fo un d in a do cu m en t. Th e bo ld ed co nt en	i N ot e Th es e ex a m pl es sh ow co m pl et e ad dr es se s as fo un d in a do cu m en t. Th e bo ld ed co nt en	00144 ROMA RM 40141 Bolo- gna (BO) 6500 Bellin- zona]Sar- ego (VI)]8407 0 - Stella Cilento - Sale- rno			91.509 -900 - Porto Alegre - RS Rio de Ja- neiro, RJ Salva- dor (BA) RIO DE JA- NEIRO Cep: 20021- 060	115035 Москв а 40000 5, Волго град	i N ot e Th es e ex a m pl es sh ow co m pl et e ad dr es se s as fo un d in a do cu m en t. Th e bo ld ed co nt en

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
					t is th e A D D RE SS 2 val ue .		t is th e AD DR ES S2 val ue .	t is th e AD DR ES S2 val ue .						t is th e AD DR ES S2 val ue .
					1234 Ma- hana St., Hono- lulu Ha- wai'i 96816		27 rue Pas- teur, Sher- brooke , Qué- bec, J1K 2Y3, Can- ada	Kaiser- straße 123, 10623 Berlin, Deutsc hland Stock- erauer straße 9, A8700 Leo- ben						Plaza de la Leal- tad, 5 28014 Ma- drid
					PO Box 10101, Bos- ton, MA 02345		4, rue du 8 Mai 1945, Van- cou- ver, BC, V6E 1R8	Zuger- berg- strass e 18, CH-64 14 Un- terae- geri						62380 Cuer- navac a, Méx- ico
					Rural Route 5, Lex- ington NE 69361		31bis, Saint- Joseph nord, 13402 Mar-	Post- fach 10 43 51, D-700 49,						28011 Ma- drid

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
-------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

seille, Stuttgart, France
 Rue du Deutschland
 Cornet
 6,
B-480
O
VERV-
IERS
 Case
 Postale
 123,
 Succursale
 Centre-Ville,
Montréal,
PQ,
 Canada
 Concours organisé en collaboration avec la commune d'**Aubenas (Ardèche)**

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
COUNTRY	آسيا	亚洲	亞洲	Azië	Asia		Asie	Afrika	Asia	アジア	유럽	América do Norte	Северная Америка	Asia
COUNTRY	أوروبا	欧洲	歐洲	Europa	Europe		Europe	Europa	Europa	欧州	남아메리카	Europa	Европа	Europa
COUNTRY	افريقيا	南美洲	南美洲	Afrika			Afrika		Africa		아프리카	Ásia	Европа	North America

i
N
o
t
e
 Au
 str
 áli
 a
 is
 ex
 tra
 ct
 ed
 as
 CO
 UN
 TR
 Y
 on
 ly.

i
N
o
t
e
 "A
 me
 ри
 ка
 "
 and
 "A
 вс
 тр
 ал
 ия
 "
 are
 ex
 tra
 ct
 ed
 as
 CO
 UN
 TR
 Y
 on
 ly.

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
COUNTRIES	البرازيل	中国	中國	Italië	Italy	افغانستان	France	Deutschland	Italia	日本	중국이	Rússia	Германия	Rusia
	كوريا الجنوبية	美国	美國	Duitsland	U.K.	ن	République	Vereinigtes Königreich	Germania	프랑스	일본은	EUA	Россия	Nicaragua
	اوزبكستان	英国	英國	U.K.	USA	آمریکا	française	Staaten von Amerika	U.K.	스	러시아	Brasil	США	Estado de Israel
	ن		英國	USA	Taiwan	کره شمالی	çaise	Amerika	USA			Portugal	Северная Корея	de Israel
	وطاجیکستان		英國	Palestijnse Nationale Autoriteit			Belgique	U.K.				Mianmar	Корея	EE UU
				Taiwan			Royaume de Belgique	Taiwan				Taiwan	США	Kosovo
							Canada						ОАЭ	Gibraltar
							Suisse							Palestina
							É.-U.							Taiwan

i Note Americas extracted as CONTEXT ONLY.

Entity Type Name	Chinese (Simplified)		Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
	Arabic	Chinese (Simplified)											
Currency [page 296]	2000 درهم	33.8 元	33 0,00	35 cents	دولار ۲۵۰	6 800	drei	€ 1.00	4 7 円	8500	\$3	80	tres
		港币五千万	萬元 1800,00	1.19 dlrs	ششصد دلار	000 DM	Dollar	0,00-1.800,00	65 円	만 달러	1 bil-hão de	мили ардов	mil cuatrocien-
		四千万	一百四十萬 1,9 mil-jard	one dollar and twenty-five cents	دولارات سبعة وعشرون مارك	68.985 FB	5€	28,5 £	7억4천700만	에	dó-lares	долларов	trocientos
		七千万	十四亿七千万 四億七千萬元	pond and twenty-five cents		300,6	40 Mil-lionen	\$ 500	파운드	240만	1 dólar	38 руб	veinte escudos
			1,19 DLRS	5.000 dollar		28,5 £	Euro	35 centesimi	엔을	15~35 달러로	153,62	к	3 pesos chilenos
			35 cent	785 to 995 dlrs		deux cent	un euro e	1.19 dlrs		1 bil-hões	de pe-sos		\$15
			100 tot 250 €	1,19 DLRS		deux yens	un euro e	25 cent		2 reais			
			12600 00000 EUR	1,19 DLRS		une dizaine de milliards de francs belges	un euro e	25 cent		R\$ 1,1 bilhão			
			tussen 32 en 33 euro	1,19 DLRS		de de mil-liards de francs	un euro e	25 cent		R			
				1,19 DLRS		de de mil-liards de francs	un euro e	25 cent		\$ 11,99			
				1,19 DLRS		de de mil-liards de francs	un euro e	25 cent		R\$ 41 mil			
				1,19 DLRS		de de mil-liards de francs	un euro e	25 cent		US \$ 0,06			
				1,19 DLRS		de de mil-liards de francs	un euro e	25 cent		centavos			
				1,19 DLRS		de de mil-liards de francs	un euro e	25 cent		de 20 a 41 dólares			
				1,19 DLRS		de de mil-liards de francs	un euro e	25 cent		dez mil euros			
			1,19 DLRS		de de mil-liards de francs	un euro e	25 cent		entre R				

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
							canadiens					\$ 1.801,37 e R		
							126000000					\$ 6.217,19		
							EUR					meio dólar		
							0.18					um euro e 3 centavos		
							EUR					um euro e 3 centavos		
							200000 USD					£ 280		
							50000000 USD					€ 2 bilhões		
							36.33					hões		
							USD					€ 20		

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
DATE	1961/3 /7	7月2日	7月2日	17 april	April 2	۸/۱۹۷۵	6 Oct	30.	25	7月26日	2012	6 de	24	15-9-96
DATE	1428/5/12	十月十七日	十月十七日	1 Januari 1999	26 November 1998	۱۳۷۴ اردیبهشت ۳ ت	17 fév 1999	September 1954	15 November 1998	2012/12/20	2012	Setembro de 2011	2012 года	15.09.96
DATE	27 يناير 15 مارس 14 نوفمبر 1995 2003	十月十七日	十月十七日	17 april	April 2	۸/۱۹۷۵	6 Oct	30.	25	7月26日	2012	6 de	24	15-9-96
DATE	1428/5/12	十月十七日	十月十七日	1 Januari 1999	26 November 1998	۱۳۷۴ اردیبهشت ۳ ت	17 fév 1999	September 1954	15 November 1998	2012/12/20	2012	Setembro de 2011	2012 года	15.09.96
DATE	27 يناير 15 مارس 14 نوفمبر 1995 2003	十月十七日	十月十七日	17 april	April 2	۸/۱۹۷۵	6 Oct	30.	25	7月26日	2012	6 de	24	15-9-96

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
Day	يوم	周一	週一	maandag	Monday	جمعه	vendredi	Montag	lunedì	火曜日	목요일	Sexta-feira	понедельник	lunes
Day	الجمعة	周六	週六	zaterdag	Monday	روز	Mardi	Mittwoch	Lunedì	にちよ (금)	(금)	Quarta-feira	к	Miércoles
Day	الأربعاء		週六	woensdag	Mon.	پنجشنبه	jeudi	Mittwoch	Giovedì	うび		Quarta-feira	пятница	viernes
Day	يوم الإثنين			dinsdag	TUES	الاثنين	et vendredi	Mo.				sábado	суббота	
Day				di			ven-dredi, samedi et dimanche	Di.						
FA											주공아			
CI											파트			
LI											롯데호			
TY											텔에서			
[page]											성수대			
296											교를			

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
FA	مطار	首都国	首都		Los									
CI	الأحساء	际机场	都		An-									
LI	مطار	浦东国	國際		geles									
TY	فاربديرو	际机场	機		Inter-									
@	مطار	中正机	場		tional									
AI	مطار	中正机	場		Airport									
R	ميلانو	場	浦											
P			東		South									
O			國際		Capitol									
RT			國際		Street									
[p			機		Heli-									
ag			場		port									
e														
297			中											
]			正											
			機											
			場											
FA	سفارة	人民公	人		Berlin									
CI	المكسي	園	民		Wall									
LI	ك	黄鹤樓	公		Dis-									
TY	المسجد	克林姆	園		ney-									
@	الأقصى	林宮	黃		land									
B	مصرف	林宮	鶴											
UI	مصرف	林宮	樓		Fort									
LD	لبنان				Knox									
G			克		Grand									
R			林		Central									
O			姆		Station									
U			林											
N			宮		Station									
D					Statue									
S					of Lib-									
[p					erty									
ag														
e														
297														
]														

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
FA	جسر	卢沟桥	盧溝橋		Cham									
CI	البحرين	重庆南	溝橋		ps-Ely-									
LI	مفرق	路	橋		sees									
TY	نابلس	王府井	重		Erie									
@	شارع	大街	慶		Canal									
PA	القدس	南路	南		Lon-									
T		忠	路		don									
H		孝			Bridge									
[p		孝												
ag		東			Times									
e		東			Square									
297		路												
]		王												
		府												
		井												
		大												
		街												
FA	مصنع	三峡工	三		San									
CI)	程	峡		Onofre									
LI	أبروكوتتر	切尔诺	工		Nu-									
TY)	贝利核	程		clear									
@	مفاعل	电站	秦		Gener-									
PL	ديمونة	小浪底	山		ating									
A	معامل	水库	核		Sta-									
N	تويوتا	电站	電		tion									
T		小	站		Three									
[p		浪			Mile Is-									
ag		底			land									
e		水												
297		庫												
]														

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
FA	ساحة	大雄宝	大		Air									
CI	السراي	殿	雄		Can-									
LI		椭圆形	寶		ada									
TY	زنزانه	椭圆形	殿		Maple									
@	غواتانام	办公室			Leaf									
S	و		橢		Loung									
U			圓		e									
B			形											
A			辦											
RE			公											
A			室											
[p														
ag														
e														
297														
]														
G	الشرق			Mid-	Gaspé-	Süd-	Mid-	東南ア	스페인	Ori-	Север	Caribe		
E	الأوسط			west	sie	deutsc	west	ジア	서쪽	ente	ная	Centro-		
O_					hland					Médio	Европ	tro-		
A	الضفة			Zui-	Ama-	zonie	Karibik	Asiat-	スラヴ	이라크	Amér-	ica	mérica	
RE	الغربية			doost-	Asie	West-	ico	五大湖	남동부	Amér-	ica	Дальн	Cáu-	
A	منطقة			Azië	du	europa	Europa	地方	동아시아	Central	Central	ий	caso	
[p	الخليج			Oost-	Sud-		dell'Es					Восто	Amér-	
ag				Europa	Est		t					к	ica	
e				Bohe-	Silicon		Boe-					Секто	ica	
297				men	Valley		mia					р Газа	Latina	
]					Moyen							Ásia-	Europa	
					-Orient							Pacíf-	del	
					Afri-							ico	Este	
					que de							Oci-	Medio	
					l'Ouest							dente	Ori-	
												Faixa	ente	
												de	Am-	
												Gaza	azonía	

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
GE_OA_REA@INDL [page 297]		华南 巴蜀 杭嘉湖	華南 巴蜀 巴蜀		North- ern Illi- nois South Florida Mid- west									
GE_OA_REA@INDL [page 298]		大中华地区 加勒比地区 加沙地带	大中华地区 加勒比地区 加沙地带		South- east Asia West- ern Eu- rope									

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
GE	بحر الشمال			Kreeftskerring		پرسیولی	delta	Mont Blanc	Tropic	富士山	지구로 부터	Atlântico	озеро Иссык-Куль	Alpes
FE	جزر بحر ایجه			Pluto		آمریکای جنوبی	Jupiter	Jupiter	del Canacro	國後島	달천강	Marte		Sierra Nevada
UA	منطقة جبل بوظهير			Lauwersmeer			Himalaya	Alpen		洋	남극해	Sumatra	Каспийское море	Cabo Cañaveral
ag				Mont Blanc			fleuve Saint-Laurent		Vesuvio			Mar do Sul da China	гора Фудзияма	Atlantis
298				Adriatische Zee			mer Ionienne		Tevere			Mar do Norte		Valle de María
]				Himalaya					Monte Bianco			península do Sinai		Anti-Illas
									Himalaya			Sistema Solar		Parque Nacional Galápagos

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
GEOGRAPHY		南北回归线	南北回归线		Mason-Dixon Tropic of Cancer									
GEOGRAPHY		地球	地球		Nephtune									
GEOGRAPHY		冥王星	冥王星		Mars									
GEOGRAPHY		北斗七星	北斗七星											

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
GEON_UFEAT_URE@WATER		峨眉山 崇明岛 珠江三角洲	峨眉山 崇明岛 珠江三角洲		Grand Canyon Mount Fuji									
GEON_UFEAT_URE@WATER		黄河 长江 西湖 日月潭	黄河 长江 長江 西湖 日月潭		Pacific Ocean Lake Michigan Volga River									

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
G					1234N									
E					/									
O					12345									
C					E									
O					LAT.									
O					12.34N									
R					LONG.									
D					012.34									
[p					W									
ag														
e					23450									
298					0S/									
]					01234									
					00W									
					12'34.5									
					N4-01									
					2'34.5									
					E6									
					3074N									
					04429									
					E									
					33 40'									
					56.14"									
					N 69									
					56'									
					20.20"									
					E									
					38° 53'									
					23"N ,									
					77° 00'									
					27"W									

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
HO-LIDAY	رأس السنة الهجرية المولد النبوي الكريسماس عيد الفصح	元宵节 中秋節 中秋	元宵節	Pasen paas- maan- dag	New Year's Day Fourth of July Martin Luther King Day	چهارشنبه ه سوری کریسماس س	Tous- saint Nouvel An Pâ- ques Réveil- lon de Noël	Weih- nachte n Tag der Ar- beit Buß- und Betttag	Venti- cinque Aprile Pas- qua Natale Festa della Re- pubbl- ica ital- iana	天皇誕 生日 こども の日 クリス マス	성탄절 신정 이드 알 아드하	Dia de Todos os Santos Fer- iado de 7 de Se- tem- bro Dia de Natal	9 мая Новый год День защит ника Отече ства Пасха	Navi- dad Epifa- nia Sem- ana Santa No- che- vieja Año Nuevo

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
LANGUAGE [page 299]	中文	英语	日本語	Een boek in het Ara-bisch . Hij spreekt Spaan-s . Het Swa-hili is een Bantu taal.	He speaks French a book written in Ara-bic .		Il parle l' espa-gnol . Un li-vre en alle-mand . Le lin-gala est une langue d'Afri-que	Eng-lisch Deutsch Portu-gie-sisch	Un li-bro in Arabo . Lui parla spag-nolo . Lo Swa-hili è una lingua bantu	日本語 ロシア語	파푸아 제어 프랑스어	USPS (na si-gla em inglês) O en-sino do portu-guês do es-panhol latim	Так меня замет ила препо-давате льниц а макед онско го языка В перев оде на русск ий некот орые англи йские выраж ения звучат очень вульга рно	el es-pañol el ruso el ale-mán el nor-uego el fran-cés

Entity Type	Chinese (Simplified)		Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
	Arabic													
LOC	جوريش	北京	北京	Amsterdam	Cairo	تهران	Honfleur	Paris	Roma	横浜	서울	Bogotá	Таганрог	Madrid
AL	خرية	上海	北京		New Delhi	واشنطن	Bruxelles	San Francisco	Parigi	上海市	로마나	Brasília	Нью-Йорк	Tel Aviv
IT	الرهوة	苏州市	台北市	Parijs	Honolulu		London	La Paz	New York		모스크바의	Jerusalém	Йорк	Miami
Y	طهران		蘇州市	New York	N.Y.C.		Prague	Freiburg	Bonn			London	Берлин	Ciudad de México
[page 299]	برلين		蘇州市	Bonn	Seville		Prague	Freiburg				London	Берлин	Ciudad de México
							San Francisco	Breisgau				New York	Санкт-Петербург	Barcelona
								Frankfurt					Санкт-Петербург	Roma
								Main					север	
								Frankfurt a.d. Oder					о-запад	
													Москва	
													ы	
													восток	
													Лондон	
													на	

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
MEASURE	خمسة أمتار مربعة ست سنتمترا ت	二百五十六公 斤 5.5 米	二百五十六公 斤 5.5 米	4.000° C	25 cubic feet	۷۴ متر	200.0 tonnes	2800 Angström	25 mq gram	60km 16 ㄹ리	17.0m 55만t	0,004 km²	5 км ² metro	3km 9 grados
MEASURE	14 ملم	5.5 米	5.5 米	800 ton	65 mph	هزارهزارگان	1.600 megawatts	58,68 mm	6m 200 tonnellate	1미터 가	106.40 0.000 km²	11 milhões de onças	12 grados centígrados	75 kilos
MEASURE				16GB	33 mpg		45 degrés	Grad Celsius	16GB 65 mph			13 km	1700 metros cúbicos de água	75 hectáreas
MEASURE				EUR 0,12 per aandeel	five cts per share		18 kilomètres	Fahrenheit	0,12 euro per azione			13 km	1700 metros cúbicos de água	siete litros
MEASURE				1,2 miljoen euro per week	per unit		30 ml	512 bits	sechshundert Kilo-gram			1700 metros cúbicos de água	1700 metros cúbicos de água	diez millas
MEASURE				van 50 tot 500 nanometer	per unit		entre 5 et 6 centimètres	entre 5 et 6 centimètres	400 euro al mese			18 cm.	18 cm.	646 toneladas
MEASURE				acht meter	per unit		de 50 à 60 eV	von 50 bis 60 Kilo-meter	von 10 Volt bis 20 Volt			2 milhões de litros	2 milhões de litros	660 megavattios
MEASURE					per unit			zwischen 5 und 6 Zentimetern	zwischen 5 und 6 Zentimetern			28 horas	28 horas	130 litros por metro cuadrado
MEASURE					per unit							280 km/h	280 km/h	de 20.00
MEASURE					per unit							30 Kg	30 Kg	de 20.00

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
								5 km/h				4 gigabytes		0 a 348.0
								33g/l				404 °C		00 hectáreas
								27 mg/Kubikmeter				500cc		eas
								5 Kilometer pro Stunde				600 mil litros		cin-cuenta kiló-metros por hora
												800m hz		
												R\$ 5 mil por mês		
												R \$197/Hora		
												US\$ 10 mil por tonelada		
												cinco kilos		
												dois acres e meio		

M	18SUH
G	6743
R	42S
S	VB
[p	7917
ag	2559
e	
299	
]	

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
MONTH	إبريل نيسان ربيع الثاني	6 月份 八月 八月 八月	6 月份 8 月份 8 月份 8 月份	januari Feb. OKT	January Feb. OCT	اردیبهش ت ماه فوريه اسفند ماه	sep-tembre mi-août	Jänner Januar März	Gen-naio Feb. OTT	7 月 七月		julho Nov. Outubro	ма рт апрель февра ль февр.	octubre Jul
NAME				t.a.v. dhr. A. Jansen	Attn: John Smith	C/O John Smith	A l'at- ten- tion de M. Robert De- larue	C.A. Mario Rossi	c/o Mario Rossi	スミス 様 気付 雅子さ ま気付		NAME _DES- IGNA- TOR Aos cuida- dos de: in "Aos Cuida- dos de Rafael de Mor- ais" CC: in "CC: Miguel Silva"	"c/o" in "c/o Mario Rossi" "Attn:" in "Attn: Mario Rossi"	

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
-------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

NI 123-45
N 6-789
@
C
A_
SI
N
[p
ag
e
300
]

i
N
ot
e
A
cu
st
o
m
Cl
ea
nsi
ng
Pa
ck
ag
e
ca
n
be
cr
ea
te
d
to
pa
rs
e
an
d
st
an
da
rdi

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
-------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

ze
NI
N/
CA
_S
IN
en
ti-
tie
s.
An
y
ex
tra
ct
ed
NI
N/
FR
_I
NS
EE
en
ti-
tie
s
ca
n
als
o
be
pa
rs
ed
an
d
st

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
							standardizing the Data Cleansing transformation of Data Services by mapping the entity to one of							

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
------------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

the UD PM (user defined pattern matching input fields. For details on using UDPM in-

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
-------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

put fields, see SAP Data Services Reference Guide

NI														
N														
@														
C														
H_														
AV														
S														
[p														
ag														
e														
300														
]														

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
-------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

NI							123 45							
N							67 890							
@							000							
FR														
_]														
N														
SE														
E														
[p														
ag														
e														
300														
]														

i
N
ot
e
A
cu
st
o
m
Cl
ea
nsi
ng
Pa
ck
ag
e
ca
n
be
cr
ea
te
d
to
pa
rs
e
an
d
st
an
da

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
-------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

rdi
ze
NI
N/
CA
_S
IN
or
NI
N/
FR
_I
NS
EE
en
ti-
tie
s.
An
y
ex
tra
ct
ed
NI
N/
CA
_S
IN
or
NI
N/
FR
_I
NS
EE

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
-------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

entities can also be parsed and standardized using the Data Cleansing Transformation of Data Set

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
							rvi ce s by m ap pi ng th e m to on e of th e UD PM (u se r def in ed pa tte rn m at ch in g) in- pu t fie							

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
------------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

Ids
.
For details on using UDP M input files, see SAP Data Services Reference Guide

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
NI									BNCM					
N									RA70A					
@I									20H50					
T_									1B					
CF														
[page 300]														
NI				51088										
N				9517										
@				69802										
N				3432										
L_														
B														
S														
N														
[page 301]														
NI											691021			
N											-10235			
@											28			
K														
R_														
R														
R														
N														
[page 301]														

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
NI					012-44									
N					-5668									
@														
U														
S_														
S														
S														
N														
[p														
ag														
e														
301														
]														

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
NOUN	العملية السياسية	新兴产业	新興產業	Nederlandse wetenschap	big-gest probleem	تشديد خشونت	pro-gres-sion	mod-erne Tech-nolo-gien	scien-ziato ital-iano	乳製品 事業 企業理念		última re-união	лесно й масси в	torneo femenino cascos
NOUN	المعلومات الوارده	高科技产品	高科技產品	inter-belan-grijke finan-ciële groep	inter-est rate mort-gage inter-est tax relief	اعتراض دانشجو	équiv-alente dével-oppe-ment dura-ble	fa-chliche Wis-sen riation-elle Ter-min-pla-nung	impor-tante grupp o fi-nan-ziario grande metro-poli			nia fixa veícu-los no-vos ano pas-sado pri-meira greve geral	сладк ие сны серый волк	azules deci-sión defini-tiva rela-ciones sino-nor-team-erica-nas verda-dera reanu-dación préstamos inmo-biliar-ios equili-brio presu-pues-tario
NOUN	الالعب الاولمبية			grote stad platen-zaak			épargn-e popu-laire							
NOUN				kinder-theater			eu-ro-péens inter-nautes expéri-menté s	Ingen-ieur- Kom-petenz						
NOUN				productiviteitscijfers										
NOUN				fab-rieks-bestel-lingen										
NOUN				speci-fieke juwe-len										
NOUN				kle-ding-stukke n										
NOUN				stu-denten										

En tit y Ty pe N a m e	Arabic	Chi- nese (Sim- pli- fied)	C hi- ne se (T ra di- tio na l)	Dutch	Eng- lish	Farsi	French	Ger- man	Italian	Japa- nese	Ko- rean	Portu- guese	Rus- sian	Span- ish
--	--------	--	--	-------	--------------	-------	--------	-------------	---------	---------------	-------------	-----------------	--------------	--------------

ouders

leraars

Entity	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ORANGE	إتشي إس بي سي	美洲银行	美 洲 銀 行	Heineken	Apple Corporation	شرکت ایران خودرو	Airbus Enron	Mercedes Benz	Fiat Fininvest	マイク ロソフト	동국제 강	Logup Consultoria Ltda.	АФК Система	Texaco Yamaha
ANIZAT	كريدی أغريکول	花旗集团	花 旗 集 团	Royal Dutch Shell	General Electric Co.	بانک کشاورز ی ایران	Northwestern Trust	Siemens	Apple Corporation	株式会社 豊田 自動織機	삼성전 자는 현대자 동차가	DuPont	Майкрософт	FIAT Corporation
ION@C	إسمنت القصيم شعاع كابينال	白云山制药股份有限公司	白 雲 山 製 藥 股 份 有 限 公 司	Apple Corporation	Apple		Banque de Montréal	Nacke und Partner	Peugeot S.A.			Brasil S.A.	Газпром	Corporación del Cobre de Chile
OMMERCIAL				Peugeot S.A.	IBM		XYZ S.A.	Texas Instruments				D&S Engenharia	Медиа МТС	la empresa SOGEMIN
[page 301]							Peugeot S.A.	Siemens AG	Otto Wolff			Apple Corporation	компания "СМВБ - Инфо	Compañía de Energía de Ceara
							Goblet Systems Enregistrée	GmbH	Kunststoffvertrieb GmbH			Peugeot S.A.	онные технологии"	
							XYZ Incorporée	Bartsch und Partner	Wolff			Apple Corporation	ЗАО "Эдем"	
							Goblet Incorporée	GmbH & Co., KG	Enregistrée	Ormecon Chemie GmbH		Peugeot S.A.	холдинг "Телекоминвест"	Автобанк-Никойл

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
							Télécom XYZ such as Télécom Inter-mo-sane	Walter de Gruyter, Inc. Greening Donald Co. Ltd.					Автомобильный Банкирский Дом Росевробанк	
							XYZ and Co. such as Ar-diton and Co.	Volkbank Hamburg Volksbank Bonn Rhein-					Банк Москвы	
							XYZ Limited such as Ad-vance Technology Limited	Dresdner Raiffeisenbank eG						

i
 N
 o
 t
 e
 A
 n
 y
 e
 x
 t
 r
 a
 c
 t
 e
 d
 O
 R
 G
 A

En tit y Ty pe N a m e	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
--	--------	-------------------------	--------------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

NI
ZA
TI
ON
/
CO
MM
ER
CI
AL
en
ti-
tie
s
ca
n
be
pa
rs
ed
an
d
st
an
da
rdi
ze
d
usi
ng
th
e
Da
ta
Qu
ali
ty
Da
ta
Cle
ans
e

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
------------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

transforming them to one of the FIRM input fields.

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ORGANIZATION		交通大学 清华 浙大	交通大學 清華 國 立 臺 灣 大 學	Universiteit van Amsterdam Vrije Universiteit Brussel Harvard University Londen Imperial College	Brown Cambridge University MIT Stanford University		Université de Bordeaux Université Pierre Marie Curie Université du Québec Ecole élémen- taire publi- que d'Amiens	Freie Universität Berlin Rheinisch- Westfälische Technische Hochschule Aachen publi- que d'Amiens	La Sapienza La Normale di Pisa Harvard University Università degli Studi di Firenze	關東學院 マギル 大學	경원대 에서 서울대 학교를	Escola Ibero- Americana de Po- lícia Universidade Kim Il- sung de Pyeong- gyang Universidade Lusófona de Hu- manidades e Tec- nologias Universidade Federal de Ciências da Saúde de Porto Alegre Harvard University	Московский государственный университет МГУ Минский лингвистический университет Массачусеттский технологический институт Escuela Elemental Rafael De Jesús	

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
-------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

Colégio Estadual Nossa Senhora de Fátima

ORGANIZATION	中央芭蕾舞团	中央芭蕾舞团	中央芭蕾舞团		Cirque du Soleil									
	上海交响乐团	上海交响乐团	上海交响乐团		Boston Symphony Orchestra									
	月之海合唱团	月之海合唱团	月之海合唱团											

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ORGANIZATION		国务院 海关总署 水利部	國務院 海關總署 水利部		Foreign Ministry Air National Guard									

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ORGANIZATION	وكالة أنباء عرب رويترز واشنطن بوست صحيفة الحياة «الرأي العام»	新华社 时代周 刊 人民日 报 人民日 報	新 華 社 時 代 週 刊 人 民 日 報		Associated Press NBC PBS									

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ORGANIZATION		中国科学院	國家科學院		American Farber Cancer Institute									
		中国科学院	中國科學院		European Space Agency									

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ORGANIZATION	الاتحاد الأوروبي	中国共产党	中國人民進步黨	UNESCO	European Community	جهه مشارکت ايران	Agence Européenne de la Sécurité Aérienne	Die Grünen	UNESCO	自民党	IAEA가	UNESCO	Ассоциация независимых центров оeconomic	Consejo de Seguridad
ORGANIZATION	العربى للمغرب العربى	联合国	聯合國	Reunited Vlaams Parlement	Benelux Greenpeace	حزب توده ايران	Sécurité Aérienne	Landesamt für Statistik	Unione Europea	ヨーク・ヤン	청와대	Agência Europeia do Ambiente	Reunited Vlaams Parlement	Consejo de Seguridad
ORGANIZATION		全国总工会	全國總工會	Fifa	United Nations	اوپک	Unesco	Greenpeace	Monti	ユベントス		Reunited Vlaams Parlement	Crusadeiro	Fronte Zapatista
ORGANIZATION					EU		Union Européenne	Partei Deutschlands	Juventus Football Club			Banco Mundial	ООН	Nacional
ORGANIZATION							Union Européenne	Zentralstelle für Agrardokumentation und -information				Organização das Nações Unidas	Министерство по налогам и сборам	Movimiento Revolucionario
ORGANIZATION												Associação das Nações do Sudeste Asiático	Тупак Амару	Cámara de Comercio
ORGANIZATION													Asociación de Hospitales Americanos de Paris	

En tit y Ty pe N a m e	Arabic	Chi- nese (Sim- pli- fied)	C hi- ne se (T ra di- tio na l)	Dutch	Eng- lish	Farsi	French	Ger- man	Italian	Japa- nese	Ko- rean	Portu- guese	Rus- sian	Span- ish
														Bene- lux
														Comu- nidad
														Eco- nó- mica
														Eu- ropea
														Unión Eu- ropea
														agen- cia de noti- cias
														Xinhua

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
ORGANIZATION		佛教	佛教		Church of Jesus Christ of Latter-day Saints									
PERSON		天主教	天主教		Church of England									

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
PERCENTAGE	%150.5 五十五 百分之五十五 خمسة وعشرون بالمائة	百分之五十五 55.3%	百分之五十五 55	220% twintig procent 15 tot 21 procent van 31% tot 25%	220% 18 procent fifty percent from 10% to 20% between 5 and 10 percent	درصد ۲۲	26.8% 6.6% de 70 à 85 %	1% + 1,23 4% 3 Prozent drei Prozent 4 1/2 Prozentpunkte 7% pro Jahr 5% / Jahr 34% / Person 2 bis 3 Prozent 2-3% 1-1,5 Prozent	220% cinquante per cento dal 10% al 20% tra il 5 e il 10 per cento	25% 10 パーセント	35%	Oi-tenta e quatro por cento cinco pontos percentuais 20% 1,43% +0,05% entre dois e cinco por cento -95,76% de 20 a 40 por cento	43 % процентов 2 процента нта	60% 53.83% % 0,8 % a 16,44 % un cinco por ciento

Entity	Chinese (Simplified)		Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish	
	Arabic	Chinese												
PE R S O N [p ag e 303]	أم علي	胡锦涛	習	Beatrix	Bill	داريوش	Maria	Mario	オバマ	노무현	Al	Валер	Roberto	
	نواف بن عبد العزيز	毛泽东	近平	Wilhelmina	Clinton	جورج بوش	Hildebrandt	Monti	菅直人	김대중	Gore e	ий	Suzanne	
	العزیز	温家宝	毛澤	Arm-gard	Clinton	ایت الله خامنه ای	Hans Peter	Obama		한승주	Georg e W. Bush	Трош ин	Prou	
	فاروق يوسف		東	Mark Rutte	Wil-liam	ای	Mayer	Angela Merkel		에게	Dilma Rousseff	ья Фотие ва	Yitzhak Rabin	
	یحیی أبو زکریا		馬英九	Barack Obama	Jefferson Clinton	خانم هدایت	Anne Marie	Wil-liam J. Clinton			Angela Merkel	Алекс ей	JUAN PABLO II	
	طارق محمد الزرقا			Angela Merkel	Mus-tafa Al-Jaziri	كمال خازری	W. J. Clinton	Vera F. Burkhardt	Giulio Terzi di Sant'A gata			Mar-cello Casal Jr.	Ивано вич Серге ев	Juan Cabal-lero Velás-quez
				Wil-liam J. Clinton	`Abd Al-Ra-haman Nudle		Mr. Bill H. Jones	Otto von Gruber	Sig.ra Ba-dano			Francisco Carlos Gomes	П.В. Шавен ков	Señor García
				Dirk Van Meer-chelen	Mary Beth Josephine		Lieut. Van Damm e		Ing. Paolo Rossi			José Gus-tavo de Souza Costa	Иван Иван	Sr. San-chez-Farrés
				Theo van Gogh	Thomas		très hon.	Jean Chrétien				de Souza Costa	Сокур ов	Sra. María José de la Garza
				Ing. Peter van den Akker	Wash-ington							Dr. Fer-nando Fa-gunde s	Джор Дж Буш-млад ший	
					Mr. Cop-per-field							Carlos Firetti, Bruno Chem-mer e Gus-	Фритц Фидле р	Татьян а Щепк

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
												tavo	ина-	
												Lobo	Купер ник	
													Сувор ова	
													Мария	
													Георг	
													иевна	

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
P	+41 22	68316	02	02 278	1-408-	۷۷۷۵۵	1-800-	+49	+39	電話 :	02-33	(11)	8(920)	(331)
H	730	616	-4	01 00	738-62	ق غ ر	555-11	1111	(06)	0566-	21-234	2872-5	284	40 41
O	5238		12	+31(0	00	۷۷۷۷۷	11	2222	322	52-212	5	502	8484	45 69
N	360-61		-9)6-388	408-73		(408)	(+49)-	0404	0	(011)2	(21)24	(+7495	Fax:
E	30		0	444 22	8-620		555-11	111-22-	+41 91	(電)	22-345	15-402) 771	(331)
[p	0		0		0		11	33333	985 88	0 8	6	2	7226	40 41
ag	617-78		0	Tel. :	0		11	33333	985 88	6 · 2	6	2	7226	40 41
e	2-266		88	+39	738-62		555-11	Telefon	55	2 4 ·	1-800-	0800-	+7(495	46 95
303	9		6-	06	00		11	01	Tel. :	4 3 2	223-4	522-15)788-9	Tel: 34
]			2-	421111	(408)		12 34	22 33	06	0	567	52	7-99	91
			25	+41 91	738-62		56 78	22 33	06					33782
			45	993 10	00		90	Tel.	421111			Celular:	телеф	00
			-5	11	1-888-		01111-1	+41 91				95363	оны	1-800-
			70	(Fax)	FLOW-		12/34	111	993 10			3463	21005	111-22
			0	12/34	ERS		/	T	11				00	22
				/	408-73		56/78	030/2	(Fax)			Fax:	или	
				56/78	8-620		/90	2 22	12/34			4803	22222	Tel: 91
				/90	0 x111		(01) 12	200	/			8118	22	111 11
				(01) 12	11 11		34 56	650	56/78			SAC:	тел.:	11
				34 56	22 22		78 90	284-61	/90			0800-	(8-051	+34
				78 90	22		+44	32	(01) 12			88870	2)-21-	111
				1-800-	11/22/		(0)		34 56			00	81-60,	22222
				555-11	33/44		1252		78 90			Tel:	49-21-	2
				11	/55		761314		1-800-			(11)	92,	111 222
				(408)	(01) 11		Tél. :	+33 1	555-11			3049-	47-88-	3333
				555-11	22 33		+33 1	11	11			0900	97	ext
				11	44 55		41 25	(408)	555-11			Tele-		1111
				555-11	(+49)-		38 15	555-11	11			fone:		Fax:
				11	111-22-		tel +32	11	11			(21)		111-22
				33333	33333		2 423	555-11	11			2688-		22
				Tele-	fon:		17 67	11	11			2629		111-22
				0111-2	2222							Tele-		22
												fone:		
												(61)		
												3411-1		
												200/0		
												1/03		

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
-------------	--------	----------------------	-----------------------	-------	---------	-------	--------	--------	---------	----------	--------	------------	---------	---------

Tels.:
(21)
2598-
9812

Tels.:
902
192
100

Tél:
+41 79
772118
0

P
RE
C
U
R
S
O
R
@
C
H
E
MI
C
AL
[p
ag
e
303
]

Cya-
nide

Di-
methyl
meth-
yl-
phos-
pho-
nate

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
PURSOR@NUCLEAR [page 303]					Cesium-137 Strontium-90									
PRODUCED [page 304]	雪碧 视窗	iPhone 6 iPhone 5 Windows 7	PlayStation iOS iPhone 4S	Windows Cheerios iPhone 4S Windows 7	Windows Angry Birds		PlayStation iPhone Instagram	iPhone Jacobs Kaffee	PlayStation iPhone 4S Windows 7	一太郎	판다클라우드 안티바이러스 이러스 LG전자 옵티머스 삼성 Z1	PlayStation iPhone 5 Windows 7	Консультант Плюс Тайд Боинг 707T	iPhone Instagram Marlboro Windows Compaq 3-5/8

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
P R O P_ M I S C [page 304]		抗日战争 八国集团首脑会议 文化大革命	抗日戰爭 八國集團首腦會議 文化大革命	Koninkrijk Paleis NASDAQ-100 Enduring Freedom	A book on the Punic Wars It is called the World Cup		Gemstar-TV EurObserver CeBIT Enduring Freedom	CeBIT Internationale Funkausstellung Europacup Olympiade	Aeroporto Malpensa NASDAQ-100 Enduring Freedom	ナスダック総合指数 万国博覧会		Barril Brent Dow Jones futuro Ibex 35 S&P 500	В финале после днего Кубка Кремля теннисистка добилась победы Заодно можно провести готовность город как Олимпиаде 2008 Между тем, согласно исследованиям ию "Аэртон", депозиты в долла	Pres-tige Guatemala- Elections Zimbabwe/N. 8

En tit y Ty pe N a m e	Arabic	Chi- nese (Sim- pli- fied)	C hi- ne se (T ra di- tio na l)	Dutch	Eng- lish	Farsi	French	Ger- man	Italian	Japa- nese	Ko- rean	Portu- guese	Rus- sian	Span- ish
													рах показа ли отриц ательн ую доход ность В годы Второ й Миров ой войны офице р был капит аном жаңда рмери и	

Entity Type	Chinese (Simplified)			Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
	Arabic													
Region	ولاية غوجرات	江苏省	江苏省	Zuid-	California	خرسان	Alsace	Hessen	Valle d'Aosta	東京都	경상도	Minas Gerais	республика	País Vasco
Region	محافظة مارب	新疆维吾尔自治区	新疆维吾尔自治区	Holland	Hawai'i	استان خرسان	Bretagne	Baden-Württemberg	Lazio	岩手県	경기도	Rio Grande do Sul	Адыгея	Canarias
Region	ولاية مارب	加利福尼亚	加利福尼亚	Noord-	Calif.	كاليفرنيا	Lorraine	Württemberg	Campania	カリフォルニア州	하와이	Sul	республика	Chiapas
Region	ولاية مينيسوتا	吾爾	吾爾	Bra-	British Columbia		Saint-Pierre-et-Miquelon	Sachsen-Anhalt	Canton Ticino			New Hampshire	Бурятия	provincia de Córdoba
Region		自治區	自治區	çao	Puerto Rico		Ontario	Califonia	Nevada			Ilhas Cook	Ямало-Ненецкий автономный округ	Cauca
Region		紐澤西	紐澤西	Wallo-	Loire			New York State	Andalusia				автoнoмный oкpyг	California
Region		西	西	nia	Bavaria			British Columbia					Московская область	
Region					Palestinian National Authority								ть	

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
Region	تلكلخ امبابة الخابورة			Vlaams-Brabant Berkshire Haute-Savoie	District of Columbia Orange County		Calvados Essonne Finistère	Landkreis Pfaffenho- fen Kreis Kelheim	provincia di Case- rta , Na- poli , Roma e Fro- sinone		금산군 종로구 에서 관악구		Волоколамский район Воскресенский район Дмитровский район	Mar- tinica Alava
Social Media	@sap-news @今日台湾 @切尔西新闻	@王襟鑫	@SCN blogs	@SCN blogs @SAP _MI-CRO-SOFT	@SCN blogs @SAP _MI-CRO-SOFT	@LaurenceDutour @René_La- ten- dresse @sap-noti- ciasbr @SCN blogs @sap-news @SAP @SAP _MI-CRO-SOFT	@Matthias_123 @RüdigerSchmitz @SCN blogs @sap-noti- ciasbr @sap-news @SAP _MI-CRO-SOFT	@SCN blogs @SAP _MI-CRO-SOFT	@sap-news @asahi _koku- sai	@HyunheeJeon @SangSangYi @갈탭 ink- orea10 @Se- cret- Gar- den_K D	@exame_com @twitter @SAP _MI-CRO-SOFT	@SCN blogs @SAP _MI-CRO-SOFT	@IsabelNevado @Ramón_Sanchez @SCN blogs @sap-noti- ciasbr @sap-news @SAP _MI-CRO-SOFT	

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
SOCIAL MEDIA TOPIC	#SAP	#祖国生日快乐 #现实版致青春#	#步步高惊心	#SAP-PRESS #SAP-project	#SAP-PRESS #SAP-project	#Ardèche #Charli-Heddo #SAP #Mo-bility	#Griechenland #Mo-bility #SAP-PRESS #SAP-project #SAP-project #Weihnachten	#SAP #SAP-PRESS #SAP-project #SAP-project #Weihnachten	#SAP #SAP-PRESS #SAP-project #SAP-project #SAP-project #Weihnachten	#서양이동양에게살을룬다 #정책金利 #JBBA #JP가10	#abc-defg #SAP-PRESS #SAP-project #SAP-project #SAP-project #SAP-project	#SAP #SAP-PRESS #SAP-project #SAP-project #SAP-project #SAP-project	#EnEl-Futuro #La_Colonia #SAP-Mobility #SAP-PRESS #SAP-project	
TICKER				SAP:N YSE NYSE: SAP	MSFT: NYSE NYSE: MSFT NAS-DAQ:A APL HPQ:N AS-DAQ	NYSE: SAP Nas-DAQ: BOBJ Nas-DAQ: US7170811035	MSFT: NYSE NYSE: MSFT NAS-DAQ:M SFT HPQ:N AS-DAQ	NYSE: SAP Nas-DAQ: BOBJ NAS-DAQ:M SFT HPQ:N AS-DAQ	NYSE: SAP Nas-DAQ: BOBJ NAS-DAQ:M SFT HPQ:N AS-DAQ	Nas-DAQ: BOBJ NAS-DAQ:A APL VSPT4 .SA PETR4 VALE5	SAP:N YSE GAZP MMB5:SBER Nas-DAQ: US7170811035	NYSE: SAP Nas-DAQ: BOBJ Nas-DAQ: US7170811035		

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
Time	الساعة الحادية عشر	8 时 3 点零 5 分	8 時 3 點 零 5 分	8:00	9:00	۴۷.۱۳	21h35	18:05:	8:00	18 時 49 分	08시 50분	15h29	9	13:45
Time	الحادية عشر	3 点零 5 分	3 點 零 5 分	21.30	9:00 a.m.	۳۰.۱	21 h 35	48	21.30	49 分	50분	(hor- ário de Brasília)	часов 15 минут	1:45 de la tarde
Time	11:30	11:30	11:30	12:00 a.m.	9:15 pm	هفت عصر	21h 35	02:00 MET	12:00 a.m.	正午	오전 11 시30분	3:25a m	5 утра	la 1.45 de la tarde
Time	الخامسة وعشر دقائق	5 分	5 分	13.30 uur	PST		21:35	16.15 Uhr	mez-zo-giorno		3:25a m	16:18	18:00	la 1.45 de la tarde
Time				half twaalf			21:35:15	2h 39	giorno			15h54		las
Time				kwart voor zes			21H00 GMT	16h 45	ore 3 e 15			meia noite		2:30 horas
Time							3		3 e un quarto			2 e quinze da madrugada		12H45 las
Time							3					19 ho-ras		las
Time							3							cinco y cuarto de la tarde
Time							3							las
Time							3							cinco y cuarto de la tarde
Time							3							las
Time							3							diez de la mañana

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
TIME [page 305]	ستين	两个月	兩個月	20 jaar	5 seconds	دقیقه ۲۰	20 ans	27 Jahre	20 anni	二時間	두 달을	déc-ada de 20	9 часов	doce horas
	25	1 小时	一個月	48 uren	1 hour.	دو ساعت	deux années	0,6 sec.	2 giorni	4 个月	1934~2010	de 17 a 25 de junho	с 5 сентября	15 minutos
	ساعة	五天	1 小 時	1 twintig jaar	35 minutes	ساعت	deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
	5		1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
	ساعات		1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos
			1 小 時	1 twintig jaar	35 minutes		deux jours	9 - 12 Mon-ate	2 due anni		김인석 (65)	de 17 a 25 de junho	с 5 сентября	15 minutos

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
			twee				juin					a um		siglo
			eeu-				2000					ano		XX
			wen				de sa-					2000-		1979-9
			jaren				medi à					2012		0
			'70				diman-							entre
			2000-1				che							1989 y
			2				entre							1991
			tussen				avril et							
			1990				juin							
			en				d'avril							
			1998				à juin							
							de mai							
							à sep-							
							tem-							
							bre							
							1896							
							entre							
							3h 30							
							et 4h							
							de 12h							
							20 à							
							6h 10							
							entre							
							1996							
							et							
							2006							
							entre							
							1991 et							
							2000							
							années							
							90							
							2005-							
							2007							

En tit y Ty pe N a m e	Arabic	Chi- nese (Sim- pli- fied)	C hi- ne se (T ra di- tio na l)	Dutch	Eng- lish	Farsi	French	Ger- man	Italian	Japa- nese	Ko- rean	Portu- guese	Rus- sian	Span- ish
--	--------	--	--	-------	--------------	-------	--------	-------------	---------	---------------	-------------	-----------------	--------------	--------------

2007/

2008

Entity Type	Chinese (Simplified)			Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
	Arabic														
TI-TL E [page 305]	لعاھل البحرين	主席	主席	voorzitter	President	وزير امور خارجہ	directeur de service	Schah	Presidente della Ar-noldo Mon-dadori S.p.A.	大統領	장관을	diretor-general da Agência In-ternacional de En-ergia	генеральный директор	el Papa	
	ووکیل وزارة الداخلية اللواء سمو الشيخ وزير العدل اللبناني الدكتو	司法部 长 总书记	總書記 總統	EU-minister-raad Ameri-kaanse minis-ter van Bui-ten-landse Zaken directeur van To-neel-groep Am-ster-dam VN-woord-voer-ster SP-minis-ter van Mobili-teit	Secretary of State Director of Mar-keting United States Attor-ney Queen of Eng-land Micro-soft CEO British Prime Minis-ter	وزیر امور خارجہ معاون بانک ملی ايران gouverneur de la Ban-que du Can-ada	ministre des Finan-ces	Papst Köni-gin Vertei-digungs-minis-terin	Presidente della Ar-noldo Mon-dadori S.p.A. ministro degli es-teri governatore della banca d'Italia	大統領 名譽教授	대통령이 상임위원장 전 새누리당 의원 보험정책연구원 부원장	Agência In-ternacional de En-ergia Atôm-ica ex-pri-meiro-minis-tro ita-liano Pre-mier do Egito chefe militar do Ha-mas CEO da Seara Foods Diretor finan-ceiro da Cesp ex-presi-	генеральный директор Агентства прикладной и регио-нальной ой полит-ики генеральный секретарь ОПЕК председате-ль Ассоци-ации комму-никационных агентов Росси-и ректор Военно-медиц-	Subco-man-dante Secre-tario de Es-tado	

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
												dente	инско	
												do	й	
												Banco	акаде	
												Central	мии	
													глава	
													Генерально	
													й	
													проку	
													ратур	
													ы	
													прези	
													дент	
													Франц	
													ии	
													адвок	
													ат экс-	
													главы	
													"ЮКО	
													Са"	

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
UR	dot_com	john_doe@usinc.com	john_doe@usinc.com	john.doe@usinc.com	john.doe@usinc.com		bruno.muri@wanadoo.fr	dot_com@usinc.com	pe-ter.schmitz@usinc.com	john.doe@usinc.com	abc_33@usinc.com	luciana_xavier@usinc.com	dot_com@usinc.com	ana_sanchez@usinc.com
EMAL		sobjects.com	@usinc.com	john.doe@usinc.com	john.doe@usinc.com		cnd@me-dia.ca		john.doe@usinc.com	somebody@usinc.com	@yna.com	FCAM-POS@ACOMINAS.COM		Du-pont/BOBJ@CMP
AGE305		support@usinc.com	jeffrey.schmitt@usinc.com	john.doe@usinc.com	john.doe@usinc.com		stephane.walton@lecho.be	Du-pont/BOBJ@CMP	luigi.rossi@chiocia.it		En-ron@CMP	presidente@planalto.gov.br		In-xight@CMP
			mary.thomas@usinc.com	Mary.Thomas@usinc.com	Mary.Thomas@usinc.com				fe-dele[mario]@chiocia.it			john.schmitt@usinc.com		
									Mary.Thomas@usinc.com			Mary.Thomas@usinc.com		
									cidha@integro-					

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
									ne-coop.org					
URI @P [page 306]	147.132.42.18	147.132.7.32	147.132.7.32	8.22.2.00.3	8.22.2.00.3	147.132.42.18	8.22.2.00.3	8.22.2.00.3	8.22.2.00.3	169.130.5.52.8	8.22.2.00.3	8.22.2.00.3	8.22.2.00.3	8.22.2.00.3
URI @P [page 306]	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com	http://www.usenet.com

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
VE	وطائرة	波音	空		Air									
HI	ستري	空中客	中		Force									
CL	بوينج	车	客		One									
E	32-707		車		Con-									
@	0				corde									
AI														
R	التومكات													
[p	أف-14													
ag	أف 35													
e														
306														
]														
VE	MER-	丰田汽	豐		blue									
HI	CEDES	车	田		1993									
CL	BENZ	凯迪拉克	汽		Volks-									
E	600	克	車		wagen									
@					Passat									
LA	همر	切诺基			1988									
N	إتش تو	吉普车			red									
D					Toyota									
[p					Camry									
ag														
e														
306														
]														

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
VE	10597				NY									
HI	لوحة 5				DGR-3									
CL	رقم				532									
E	586				CA									
@	اللوحه :				AVC36									
LI-	أم ق				7									
C	/													
E	رقمها /													
N	00035													
SE	4													
[p														
ag														
e														
306														
]														
VE	124875				1G1JF2									
HI	شاسيه 1				7W8GJ									
CL	VIN:				178227									
E	2MEF													
@	M74W													
VI	5WX67													
N	5387													
[p	651014													
ag	2													
e														
306	651014													
]	ورقم 2													
	الموتور													

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
VE	U.S.S	泰坦尼	永		USS									
HI	Trux-	克	豐		Cole									
CL	tun	永丰舰	艦		USS									
E	كول		遼		Con-									
@			寧		stitu-									
W			號		tion									
AT			航											
ER			空											
[p			空											
ag			母											
e			艦											
306														
]														
W		炭疽武	炭		An-									
E		器	疽		thrax									
A			武											
P			器		ricin									
O														
N														
@														
BI														
O														
LO														
GI														
C														
AL														
[p														
ag														
e														
307														
]														

Entity Type	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
Wikipedia		芥子毒气	芥子毒氣		VX tabun									
Wikipedia		TNT 炸药	TNT 炸藥		Molotov Dynamite									

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
Wikipedia@PURL [page 307]		长崎原 子弹	廣 島 原 子 彈		A- bomb pluto- nium									
Wikipedia@PURL [page 307]		响尾蛇 导弹 赤链蛇 飞弹	赤 鏈 蛇 飛 彈		Al-Sa- moud 2 Silk- worm									

Entity Type Name	Arabic	Chinese (Simplified)	Chinese (Traditional)	Dutch	English	Farsi	French	German	Italian	Japanese	Korean	Portuguese	Russian	Spanish
WEAPON@SHOOTING [page 307]		卡宾枪 来福枪	卡 賓 槍 來 福 槍		AK-47 AKM									
YEAR [page 307]	2005 年 一九九 四年	2005 年 一九九 四年	2005 年 一九九 四年	2012 '48 jaar 300 v.Chr	2001 '63 1998 A.D. 200 BC 2525 C.E.	سال ۸۱ 2007		'99 58 vor Chris- tus 200 v.Chr. 3 n. Chr.	2012 '48 300 A.C.	2012 年 平成 23 年	2012 년 A.D. 2000	2012 ano de 680 450 a.C.	1998 год 2000 год	1982 '68 2013

5.3 Definitions of Entity Types

These are the names and definitions of the predefined entity types.

5.3.1 ADDRESS1

Corresponds to the first part of a postal address. At least two pieces of the following address information must be adjacent to be extracted as ADDRESS1: Building name, House number, Street name, Apartment number, PO Box and number, Neighborhood. The format is typical of the country (or countries) where the language is spoken

5.3.2 ADDRESS2

Corresponds to the second part of a postal address. At least two pieces of the following address information must be adjacent to be extracted as ADDRESS2: Postal code, City, State/province/department. Country names are always extracted as COUNTRY, even in an address context.

5.3.3 CONTINENT

The name of any continent.

5.3.4 COUNTRY

The name, or the abbreviated name of a country.

5.3.5 CURRENCY

A quantity of world currency, or a range of amounts of currency.

5.3.6 DATE

A date is minimally composed of a number and month.

5.3.7 DAY

A day of the week, including abbreviations.

5.3.8 FACILITY

A man-made structure.

5.3.9 FACILITY@AIRPORT

The name of a primarily man-made or man-maintained structure whose primary use is as an air transportation terminal.

5.3.10 FACILITY@BUILDGROUNDS

The name of an architectural or civil engineering structure, or an outdoor space that is mainly man-made or man-maintained. There is no distinction with respect to their function, they could be civil or military facilities, they could be used for work or entertainment, or they could be monuments.

5.3.11 FACILITY@PATH

The name of a primarily man-made or man-maintained structure that allows fluids, energy, persons, animals, or vehicles to pass from one location to another.

5.3.12 FACILITY@PLANT

The name of a facility composed of one or more buildings used for industrial purposes.

5.3.13 FACILITY@SUBAREA

The name of a portion of a facility, typically an architectural one, that is able to contain people, animals, or objects.

5.3.14 GEO_AREA

A geographical area larger than a city and consisting of a significant land mass, such as a group of countries.

5.3.15 GEO_AREA@DOMESTIC

The name of a location that does not cross national boundaries.

5.3.16 GEO_AREA@INTL

The name of a location that crosses national boundaries.

5.3.17 GEO_FEATURE

A geographical location that does not constitute a political entity and is not a CONTINENT, GEO_AREA, COUNTRY, REGION, or LOCALITY.

5.3.18 GEO_FEATURE@BOUNDARY

The name of a location such as a border.

5.3.19 GEO_FEATURE@CELESTIAL

The name of an astronomical location that is outside of the boundaries of the Earth

5.3.20 GEO_FEATURE@LAND

The name of a location that is a geologically or ecosystemically designed, nonartificial location.

5.3.21 GEO_FEATURE@WATER

The name of a location that is a body of water.

5.3.22 GEOCOORD

Geographic coordinates, of various formats.

5.3.23 HOLIDAY

Holidays when banks and businesses are closed (bank holidays) in countries where that language is spoken. Spelled out dates that are also names of holidays are extracted as HOLIDAY (like "Fourth of July") whereas numeric dates that coincide with a holiday are extracted as DATE (like, "4th of July").

5.3.24 LANGUAGE

The name of a language.

5.3.25 LOCALITY

Name of a city.

5.3.26 MEASURE

A measure expression, or a range expression denoting measure.

5.3.27 MGRS

Military Grid Reference System coordinates, of various formats.

5.3.28 MONTH

The name of a month of the year, including abbreviations.

5.3.29 NAME_DESIGNATOR

A designator that appears before a person's name.

5.3.30 NIN@CA_SIN

A Canadian Social Identification number.

i Note

A custom Cleansing Package can be created to parse and standardize `NIN@CA_SIN` entities. Any extracted `NIN@CA_SIN` entities can also be parsed and standardized using the Data Cleanse transform of Data Services by mapping them to one of the UDPM (user defined pattern matching) input fields.

For details on using UDPM input fields, see *SAP Data Services Reference Guide*

5.3.31 NIN@CH_AVS

A Swiss national identification number: Assicurazione Vecchiaia e Superstiti.

i Note

A custom Cleansing Package can be created to parse and standardize `NIN/IT_CF` or `NIN/IT_AVS` entities. Any extracted `NIN/IT_CF` or `NIN/IT_AVS` entities can also be parsed and standardized using the Data Cleanse transform of Data Services by mapping them to one of the UDPM (user defined pattern matching) input fields.

For details on using UDPM input fields, see *SAP Data Services Reference Guide*

5.3.32 NIN@FR_INSEE

A French national identification number: Institut National de la Statistique et des Études Économiques.

i Note

A custom Cleansing Package can be created to parse and standardize `NIN/FR_INSEE` entities. Any extracted `NIN/FR_INSEE` entities can also be parsed and standardized using the Data Cleanse transform of Data Services by mapping them to one of the UDPM (user defined pattern matching) input fields.

For details on using UDPM input fields, see *SAP Data Services Reference Guide*

5.3.33 NIN@IT_CF

An Italian national identification number: Codice Fiscale.

i Note

A custom Cleansing Package can be created to parse and standardize `NIN/IT_CF` or `NIN/IT_AVS` entities. Any extracted `NIN/IT_CF` or `NIN/IT_AVS` entities can also be parsed and standardized using the Data

Cleanse transform of Data Services by mapping them to one of the UDPM (user defined pattern matching) input fields.

For details on using UDPM input fields, see *SAP Data Services Reference Guide*

5.3.34 NIN@KR_RRN

A Korean resident registration number (주민등록번호).

5.3.35 NIN@NL_BSN

A Dutch national identification number: Burgerservicenummer.

i Note

The extractions are not validated against the Elfproef (eleven test)

5.3.36 NIN@US_SSN

National identifier used in the USA : Social Security Number. Any extracted NIN/US_SSN entities can be parsed and standardized using the Data Cleanse transform of Data Services by mapping them to one of the SSN input fields.

5.3.37 NOUN_GROUP

A NOUN_GROUP is any common noun sequence consisting of one or more related nouns, or modifier(s) plus noun(s), which are not identified as name, measure, or identifier. NOUN_GROUP is extracted by all language modules except Greek.

5.3.38 ORGANIZATION@COMMERCIAL

The name of commercial organizations, such as major companies or corporations. Any extracted ORGANIZATION/COMMERCIAL entities can be parsed and standardized using the Data Quality Data Cleanse transform by mapping them to one of the FIRM input fields.

5.3.39 ORGANIZATION@EDUCATIONAL

The name of an institution focused primarily on education.

5.3.40 ORGANIZATION@ENTERTAINMENT

The name of any organization, whether nonprofit or for profit, whose main interest is in the production of performing arts material or events: dance, music, opera, theatre, magic, spoken word, circus arts, and musical theatre.

5.3.41 ORGANIZATION@GOVERNMENT

The name of an organization related to government, politics, or the state.

5.3.42 ORGANIZATION@MEDIA

The name of any organization, whether nonprofit or for profit, at any level, whose main interest is in the production of publishing and/or radio-, TV-, cable-, satellite-, or web-broadcasting material.

5.3.43 ORGANIZATION@MEDICALSCIENCE

The name of an organization focused on medical care or research.

5.3.44 ORGANIZATION@RELIGIOUS

The name of an organization focused on religion.

5.3.45 ORGANIZATION@SPORTS

The name of an organization focused on sports.

5.3.46 ORGANIZATION@OTHER

The name of any organization that does not fit into a more specific subtype, including groupings of geopolitical entities that can function as political entities.

5.3.47 PEOPLE

A name referring to an identifiable group of people based on nationality. In certain languages, ethnicity, region, or religion are also recognized.

5.3.48 PERCENT

A percent expression, or a range expression denoting percentage.

5.3.49 PERSON

An individual specified by name. A variety of forms are identified. Given and family names that occur by themselves are extracted as PERSON as long as they are not ambiguous with common names, with the exception of famous cases, such as Bush.

5.3.50 PHONE

A telephone number based on the patterns used in the USA and internationally.

5.3.51 PRECURSOR@CHEMICAL

The name of a chemical substance that has been officially identified as used in the manufacture of chemical weapons.

5.3.52 PRECURSOR@NUCLEAR

The name of a chemical substance that has been officially identified as used in the manufacture of nuclear weapons.

5.3.53 PRODUCT

A product name, optionally preceded by a company name.

5.3.54 PROP_MISC

A proper name that does not fall into any of the entity types specified by the other entities.

5.3.55 REGION@MAJOR

The major administrative division of a country, such as a state of the US, including abbreviations, a province or territory of Canada, the regions of England, and so on. This list also includes names of geo-political entities for which conventional labels do not apply, such as disputed territories or territories that have not been internationally recognized.

5.3.56 REGION@MINOR

The name of a county, prefecture, district, or analogous geographical division or governmental unit.

5.3.57 SOCIAL_MEDIA@ID_TWITTER

A Twitter handle or ID starting with "@".

5.3.58 SOCIAL_MEDIA@TOPIC_TWITTER

A Twitter topic name: a character sequence starting with "#".

5.3.59 TICKER

A company stock ticker symbol used on the stock exchange. If TICKER entities are found adjacent to index or market names, these are included in the span.

5.3.60 TIME

A clock time or a time expression.

5.3.61 TIME_PERIOD

A measure of time duration or an expression denoting a range of measures of time.

5.3.62 TITLE

The name of a position in government, business, or some other organization.

5.3.63 URI@EMAIL

An e-mail address, including Lotus Notes addresses.

5.3.64 URI@IP

An Internet Protocol address.

5.3.65 URI@URL

A uniform resource locator, or web address.

5.3.66 VEHICLE@AIR

An air vehicle, such as an airplane or helicopter.

5.3.67 VEHICLE@LAND

A land vehicle, including the color, year, model and make of the vehicle.

5.3.68 VEHICLE@LICENSE

An alphanumeric sequence that conforms to the US and Canadian license plate formats, when preceded by a state abbreviation.

5.3.69 VEHICLE@VIN

A Vehicle Identification Number (VIN), which always consists of 17 characters. Each position in the VIN has a particular meaning, designating among other things county code, manufacturer code, equipment code, serial number, and so on.

5.3.70 VEHICLE@WATER

A water vehicle.

5.3.71 WEAPON@BIOLOGICAL

The name of some bacteria, virus, fungus, natural toxin, or disease that has been officially identified as used to harm humans, plants (crops), and animals, or as a potential biological threat. This also extracts entities that describe the means for the dispersal of any of these weapons.

5.3.72 WEAPON@CHEMICAL

The name of a chemical substance that has been officially identified as used to harm humans, plants (crops), and animals, or as a potential chemical threat. This also extracts entities that describe the means of the dispersal of any of these weapons.

5.3.73 WEAPON@EXPLODING

The name of a substance or weapon that is designed or used to explode for the purpose of causing damage.

5.3.74 WEAPON@NUCLEAR

The name of a weapon that has been officially identified as used to harm humans, plants (crops), and animals through the dispersal of radiological or nuclear energies, or has been identified as a potential nuclear threat.

5.3.75 WEAPON@PROJECTILE

The name of a weapon that is designed or used to be projected at great speed for the purpose of causing damage.

5.3.76 WEAPON@SHOOTING

The name of a weapon that is designed or used to send projectile objects at great speed for the purpose of causing damage.

5.3.77 YEAR

A year identifier.

5.4 Entity Normalization

Normalization is the process of standardizing a span of text to conform to a predefined standard. For instance, Text Analysis normalizes numeric entities of the type CURRENCY into standard formats defined by the International Standards Organization.

i Note

CURRENCY, DATE, MEASURE, PERCENT, TIME_PERIOD, and YEAR are the entity types currently supported. Entity normalization is included in the English language for these six types.

German covers DATE, MEASURE, TIME_PERIOD, and YEAR.

French covers only MEASURE.

i Note

Entity normalization is not on by default. To activate it, please refer to "Entity Normalization" in the *SAP HANA Text Analysis Developer Guide*.

Currency

Text Analysis currency normalizer converts surface currency forms, e.g. **\$15.75**, to a standard representation incorporating the three-character currency codes of ISO 4217:2015. The output format is <Amount><Space><ISO 4217:2015 three character-currency code>, as in **15.75 USD**.

Some currency expressions, such as peso, colon, or dinar, require country indication for the normalizer to work correctly. If no country information is found along these currency names, default normalization occurs. For example, **dollar** defaults to USD, **franc** to CHF, **pound** to GBP, or **rupee** to INR.

Examples of the English currency normalizer:

Input	Output
\$100	100 USD
£5 GBP	5 GBP
c\$5.27	5.27 CAD
HK\$ 400	400 HKD
850 MXN	850 MXN
m\$3.32	3.32 MXN
2,749.57 Canadian dollars	2749.57 CAD

Input	Output
.99 Omani Rials	.99 OMR
114 cents	1.14 USD
five pennies	.05 USD
25 cts	.25 USD
37.5¢	.375 USD
2 centimes	.02 CHF
CAD 80-85	80 CAD - 85 CAD
14.2 billion yen	14200000000 JPY
\$157M	157000000 USD
a hundred Turkish lira	100 TRY
30-40 dollars	30 USD - 40 USD
between 600 and 700 Mexican pesos	600 MXN - 700 MXN

If the CURRENCY entity cannot be normalized, it is left unchanged.

Date

The DATE normalizer converts date expressions, e.g. **1/13/01** to a standard representation. The output format is <YYYY>-<MM>-<DD>, as in **2001-01-13**. DATE normalization follows the TIMEX3 standard which in turn is based on the extended ISO 8601 format.

Partial DATE expressions are normalized if they either contain day and month or month and year.

Examples of the English date normalizer:

Input	Output
15.2.09	2009-02-15
March 7th, 2016	2016-03-07
15April	XXXX-04-15
October, 1985	1985-10-XX

Examples of the German date normalizer:

Input	Output
15.2.09	2009-02-15
7. März 2016	2016-03-07
15. April	XXXX-04-15
Juli 1776	1776-07-XX

Measure

The MEASURE normalizer converts measurement expressions, e.g. **34 CM** and **0.5 kilos**, to a standard representation. The output format is <Value><space><Unit>, as in **34 cm** and **.5 kg**. The units of measurement are normalized according to the International System of Units from the National Institute of Standards and Technology.

Examples of the English measure normalizer:

Input	Output
5 inches	5 in
35 kilos	35 kg
90 km/hour	90 km/h
from 3 to 5 mg	3 mg - 5 mg
from 5 feet 5 inches to 6 feet	5 ft 5 in - 6 ft
3mm x 5mm	3 mm x 5 mm
five hundred meters	500 m
>= 1500 to < 2000 mg	>= 1500 mg - < 2000 mg
half a square foot	.5 ft ²
1.4mg per tablet	1.4 mg per tablet

Examples of the German measure normalizer:

Input	Output
5 Zentimeter	5 cm
35 Kilogramm	35 kg

Input	Output
90 Kilometer pro Stunde	90 km/h

Examples of the French measure normalizer:

Input	Output
5 centimètres	5 cm
35 kilos	35 kg
90 km/heure	90 km/h

Percent

The PERCENT normalizer converts percent expressions, e.g. **15%** and **five hundred percent**, to a standard representation. The output format is <Amount>%, as in **15.25%**.

Examples of the English percent normalizer:

Input	Output
21%	21%
14.5%	14.5%
157 PERCENT	157%
-212.402 Perc.	-212.402%
minus twelve percent	-12%
positive 22.37 Perc.	+22.37%
Zero Percent	0%
thirteen percentage points	13%
seventeen pc.	17%
14 4/5%	14.8%
one-half of a percentage point	.5%
a thousand percent	1000%
forty-seven to forty-nine percent	47-49%

Input	Output
2.5-7%	2.5-7%

If the PERCENT entity cannot be normalized, it is left unchanged.

Time Period

TIME_PERIOD entities denoting duration are normalized. Entities expressing duration in years, months, weeks, days, hours, minutes, or seconds, follow the ISO 8601 standard for normalization:

- PnnW (P4W = period of 4 weeks)
- PnnYnnMnnD (P2Y6M20D = period of 2 years 6 months 20 days)
- PTnnHnnMnnS (PT3H22M12S = period of time of 3 hours 22 minutes 12 seconds)

i Note

Date ranges (from Sept 15 to Oct 10), day ranges (Monday-Friday), time ranges (8-5 pm), and ages (83 years old) are not normalized at this point.

Examples of the English time period normalizer:

Input	Output
4 weeks	P4W
4 years, 3 months and 5 days	P4Y3M5D
5 minutes	PT5M

Examples of the German time period:

Input	Output
vier Wochen	P4W
vier Jahre, drei Monate und elf Sekunden	P4Y3MT11S

Year

The YEAR normalizer converts year expressions, e.g. '67 to a standard representation. The output format is <YYYY>, as in 1967.

In the case of years in two-digit representation, e.g. '87, '03, without any common-era (CE) or before-common-era (BCE) suffix, the system assumes 10 to 99 to refer to years in the 20th century, and 00/01 to 09/19 to

represent years in the 21st century. Because of this, the mere fact of standard representation should not be interpreted as a guarantee of accuracy.

Examples of the English year normalizer:

Input	Output
1973	1973
1754 A.D	1754
'99	1999
'02	2002

Examples of the German year normalizer:

Input	Output
1973	1973
'80	1980

6 Fact Extraction

Fact extraction includes detection of sentiments expressed about something, public-sector-specific information and enterprise facts.

6.1 Sentiment Analysis Fact Extraction

Sentiment analysis content includes a set of entity types and rules that address requirements for extracting customer sentiments, requests, emoticons, and profanities.

You can use this content to retrieve specific information about your customers' needs and perceptions when processing and analyzing text.

Sentiment analysis content is included in and supports these language modules:

- Arabic
- Chinese (simplified)
- Chinese (traditional)
- Dutch - Emoticons and profanities only
- English
- French
- German
- Italian
- Portuguese
- Russian
- Spanish

Fact extraction involves complex linguistic analysis and pattern matching that includes processing parts of speech, syntactic patterns, negation, and so on, to identify the patterns to be extracted.

The extraction output includes the identified patterns and information about each extraction, including the type of information extracted (either sentiment or request).

Sentiment analysis content includes the following rule sets and dictionaries:

Rule Set Description

- Sentiment
Extracts information about sentiments and problems, including the strength of the sentiment, such as strong or weak
 - Emoticon
Extracts information about sentiments expressed by emoticons or emojis, which generally convey the user's feelings towards the whole utterance and not a particular topic within it
-

Rule Set Description

- Request
Extracts general requests made by the customer, including requests to be contacted
-

When analyzing and extracting data, the rule sets consider a statement that expresses a customer sentiment or a request as a disposition. Dispositions are further divided into stances and topics:

- For sentiment dispositions, the stance represents the category or type of sentiment, for example, a strong positive sentiment or a strong negative sentiment.
- For a request disposition, the stance represents the type of request: general or contact.
- Emoticon rule sets extract only stances, i.e. sentiments, without identifying topics.
- A topic represents what the sentiment or the request is about.

Each extracted disposition typically includes:

- At least one stance.
- An optional topic; multiple topics are allowed within one disposition.

i Note

For details about using sentiment analysis content to enhance extraction, refer to the *SAP Data Services Text Data Processing Extraction Customization Guide*.

For a full list of entity and subentity names, see [Entity Type Names and Subentity Type Names For Fact Extraction \[page 378\]](#) and note those whose name begins with "VoC" (Voice of the Customer).

Related Information

[Predefined Entity Types by Language \[page 208\]](#)

[Sentiment Extraction \[page 315\]](#)

[Emoticon Extraction \[page 329\]](#)

[Request Extraction \[page 340\]](#)

[Profanity Extraction \[page 346\]](#)

6.1.1 Sentiment Extraction

Sentiment rules are designed to extract information about someone's feelings about something. The rules extract patterns that express customer feelings about concepts, places, actions, items, and so forth; for example, a product, company, service, or person.

The rules categorize extracted sentiments into the following types of stances:

Type of Sentiment Stance	Description
Strong positive sentiment	A strong positive opinion, such as "great" or "excellent"
Weak positive sentiment	A weak positive opinion, such as "nice" or "fine"
Neutral sentiment	An opinion that expresses ambivalence, such as "OK" or "acceptable"
Weak negative sentiment	A weak negative opinion, such as "bad" or "dislike"
Strong negative sentiment	A strong negative opinion, such as "hate" or "terrible"
Minor problem	An opinion describing an impediment the customer can work around, such as "too slow" or "buggy"
Major problem	An opinion describing an impediment the customer cannot work around, such as "incomprehensible" or "defective"

In addition to a stance, an extracted sentiment optionally includes one topic—the subject of the sentiment. The topic answers the question "What is it that the customer is expressing their feelings about?"

6.1.1.1 Arabic Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in Arabic.

Strong Positive Sentiment

- كانت لوحة المفاتيح ممتازة

StrongPositiveSentiment	ممتازة
Topic	لوحة المفاتيح

Weak Positive Sentiment

- كانت لوحة المفاتيح جيدة

WeakPositiveSentiment	معقدة
Topic	لوحة المفاتيح

Neutral Sentiment

- لوحة المفاتيح كانت لا بأس بها

NeutralSentiment	لا بأس بها
Topic	لوحة المفاتيح

Weak Negative Sentiment

- كانت لوحة المفاتيح سيئة

WeakNegativeSentiment	سيئة
Topic	لوحة المفاتيح

Strong Negative Sentiment

- كانت لوحة المفاتيح فظيعة

StrongNegativeSentiment	فظيعة
Topic	لوحة المفاتيح

Minor Problem

- كانت لوحة المفاتيح معقدة

MinorProblem	معقدة
Topic	لوحة المفاتيح

Major Problem

- كانت لوحة المفاتيح عطلانة

MajorProblem	عطلانة
--------------	--------

6.1.1.2 Chinese (Simplified) Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in Simplified Chinese.

Strong Positive Sentiment

- 我爱这辆车。
我[StrongPositiveSentiment] 爱[/StrongPositiveSentiment] 这辆[Topic] 车[/Topic] 。

Weak Positive Sentiment

- 屏幕质量不错。
[Topic] 屏幕质量[/Topic] [WeakPositiveSentiment] 不错[/WeakPositiveSentiment] 。

Neutral Sentiment

- 我们遵守这个约定。
我们[NeutralSentiment] 遵守[/NeutralSentiment] 这个[Topic] 约定 [/Topic] 。

Weak Negative Sentiment

- 我们小看了他。
我们[WeakNegativeSentiment] 小看[/WeakNegativeSentiment] 了他。

Strong Negative Sentiment

- 他非常贪婪。
他 [StrongNegativeSentiment] 非常贪婪 [/StrongNegativeSentiment] 。

Minor Problem

- 方式有些复杂。
[Topic] 方式 [/Topic] 有些 [MinorProblem] 复杂 [/MinorProblem]。

Major Problem

- 电脑突然死机了。
[Topic] 电脑 [/Topic] 突然 [MajorProblem] 死机 [/MajorProblem] 了。

6.1.1.3 Chinese (Traditional) Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in Traditional Chinese.

- 我愛這輛車。
我 [StrongPositiveSentiment] 愛 [/StrongPositiveSentiment] 這輛 [Topic] 車 [/Topic] 。

Weak Positive Sentiment

- 螢幕品質不錯。
[Topic] 螢幕品質 [/Topic] [WeakPositiveSentiment] 不錯 [/WeakPositiveSentiment] 。

Neutral Sentiment

- 我們遵守這個約定。
我們 [NeutralSentiment] 遵守 [/NeutralSentiment] 這個 [Topic] 約定 [/Topic] 。

Weak Negative Sentiment

- 我們小看了他。
我們 [WeakNegativeSentiment] 小看 [/WeakNegativeSentiment] 了他。

Strong Negative Sentiment

- 他非常貪婪。
他 [StrongNegativeSentiment] 非常貪婪 [/StrongNegativeSentiment]。

Minor Problem

- 方式有些複雜。
[Topic] 方式 [/Topic] 有些 [MinorProblem] 複雜 [/MinorProblem]。

Major Problem

- 電腦突然出故障了。
[Topic] 電腦 [/Topic] 突然出 [MajorProblem] 故障 [/MajorProblem] 了。

6.1.1.4 English Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in English.

Strong Positive Sentiment

- I'm so excited about this car!
I'm so [StrongPositiveSentiment] excited [/StrongPositiveSentiment] about this [Topic] car
[/Topic].

Weak Positive Sentiment

- The screen is nice with a good picture quality.
The [Topic] screen [/Topic] is [WeakPositiveSentiment] nice [/WeakPositiveSentiment] with
a [WeakPositiveSentiment] good [/WeakPositiveSentiment] [Topic] picture quality [/Topic].

Neutral Sentiment

- I don't love your software.

I don't[NeutralSentiment] love [/NeutralSentiment] your [Topic] software [/Topic].

Weak Negative Sentiment

- The coffee was disappointing.
The [Topic] coffee [/Topic] was [WeakNegativeSentiment] disappointing [/WeakNegativeSentiment].

Strong Negative Sentiment

- My new laptop is absolutely terrible.
My [Topic] new laptop [/Topic] is absolutely [StrongNegativeSentiment] terrible [/StrongNegativeSentiment].

Minor Problem

- Most flights are overbooked.
Most [Topic] flights [/Topic] are [MinorProblem] overbooked [/MinorProblem].

Major Problem

- It seems to be defective.
It seems to be [MajorProblem] defective [/MajorProblem].

6.1.1.5 French Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in French.

Strong Positive Sentiment

- Les repas y sont excellents.
Les [Topic] repas [/Topic] y sont [StrongPositiveSentiment] excellents [/StrongPositiveSentiment].

Weak Positive Sentiment

- Nous sommes contents de notre nouvelle voiture.
Nous sommes [WeakPositiveSentiment] contents [/WeakPositiveSentiment] de notre [Topic] nouvelle voiture [/Topic].

Neutral Sentiment

- Une voiture relativement acceptable.
Une [Topic] voiture [/Topic] relativement [NeutralSentiment] acceptable [/NeutralSentiment].

Weak Negative Sentiment

- Un serveur assez impoli.
Un [Topic] serveur [/Topic] assez [WeakNegativeSentiment] impoli [/WeakNegativeSentiment].

Strong Negative Sentiment

- Le tirage photo me déçoit beaucoup.
Le [Topic] tirage photo [/Topic] me [StrongNegativeSentiment] déçoit [/StrongNegativeSentiment] beaucoup.

Minor Problem

- Le problème se situe dans le moteur.
Le [MinorProblem] problème [/MinorProblem] se situe dans le [Topic] moteur [/Topic].

Major Problem

- Contrairement à ce que l'emballage indique, ça ne fonctionne pas.
Contrairement à ce que l'emballage indique, ça ne [MajorProblem] fonctionne pas [/MajorProblem].

6.1.1.6 German Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in German.

Strong Positive Sentiment

- Der Fernseher ist klasse.
Der [Topic]Fernseher[/Topic] ist [StrongPositiveSentiment] klasse[/StrongPositiveSentiment].

Weak Positive Sentiment

- Ich mag das Radio.
Ich [WeakPositiveSentiment] mag [/WeakPositiveSentiment] das [Topic] Radio[/Topic].

Neutral Sentiment

- Ich finde Ihre Produkte akzeptabel.
Ich finde Ihre [Topic]Produkte[/Topic] [NeutralSentiment] akzeptabel [/NeutralSentiment].

Weak Negative Sentiment

- Der Hauptbahnhof ist nicht schön.
Der [Topic]Hauptbahnhof[/Topic] ist nicht [WeakNegativeSentiment] schön[/WeakNegativeSentiment].

Strong Negative Sentiment

- Der Service war furchtbar.
Der [Topic] Service[/Topic] war [StrongNegativeSentiment] furchtbar [/StrongNegativeSentiment].

Minor Problem

- Es sieht so aus, als ob es meinem Computer schadet.
Es sieht so aus, als ob es meinem [Topic]Computer [/Topic] [MinorProblem] schadet [/MinorProblem].

Major Problem

- Die Installierung hat meinen Computer kaputt gemacht!.
Die Installierung hat meinen [Topic]Computer [/Topic] [MajorProblem] kaputt gemacht [MajorProblem]!

6.1.1.7 Italian Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in Italian.

Strong Positive Sentiment

- Eccellente carrozzeria.
[StrongPositiveSentiment] Eccellente [/StrongPositiveSentiment] [Topic] carrozzeria [/Topic].

Weak Positive Sentiment

- Non sono pentito di aver acquistato questa vettura.
Non sono [WeakPositiveSentiment]pentito [/WeakPositiveSentiment] di [topic] aver acquistato questa vettura [/topic].

Neutral Sentiment

- Questo prodotto mi lascia indifferente.
Questo [topic]prodotto [/topic] mi lascia [NeutralSentiment]indifferente [/NeutralSentiment].

Weak Negative Sentiment

- Il ristorante era brutto.
Il [topic] ristorante [/topic] era [WeakNegativeSentiment] brutto [/WeakNegativeSentiment].

Strong Negative Sentiment

- Il dispositivo è terribile.
Il [topic] dispositivo [/topic] è [StrongNegativeSentiment] terribile [/StrongNegativeSentiment].

Minor Problem

- Quest'auto ha un odore strano.
Quest' [topic] auto [/topic] ha un [MinorProblem] odore strano [/MinorProblem].

Major Problem

- La tastiera si è rotta.
La [topic] tastiera [/topic] si è [MajorProblem] rotta [/MajorProblem].

6.1.1.8 Portuguese Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in Portuguese.

Strong Positive Sentiment

- O livro me fascina!
O [Topic] livro [/Topic] me [StrongPositiveSentiment] fascina [/StrongPositiveSentiment]!

Weak Positive Sentiment

- O livro é bom.
O [Topic] livro [/Topic] é [WeakPositiveSentiment] bom [/WeakPositiveSentiment].

Neutral Sentiment

- O programa é mais ou menos.
O [Topic] programa[/Topic] é [NeutralSentiment] mais ou menos[/NeutralSentiment].

Weak Negative Sentiment

- Me irrita a espera pela manutenção.
Me [WeakNegativeSentiment] irrita[/WeakNegativeSentiment] a [Topic] espera pela manutenção[/Topic].

Strong Negative Sentiment

- Odeio o livro!
[StrongNegativeSentiment] Odeio[/StrongNegativeSentiment] o [Topic] livro[/Topic]!

Minor Problem

- A linguagem é difícil.
A [Topic] linguagem[/Topic] é [MinorProblem] difícil[/MinorProblem]

Major Problem

- O teclado quebrou.
O [Topic] teclado[/Topic] [MajorProblem] quebrou[/MajorProblem].

6.1.1.9 Russian Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in Russian.

Strong Positive Sentiment

- Купил и не жалею, улётный планшет.
Купил и не жалею, [StrongPositiveSentiment] улётный[/StrongPositiveSentiment] [Topic] планшет[/Topic].

Weak Positive Sentiment

- Объективом доволен.
[Topic]Объективом[/Topic] [WeakPositiveSentiment]доволен[/WeakPositiveSentiment].

Neutral Sentiment

- Что мне мешает нормально работать.
Что мне мешает [NeutralSentiment]нормально[/NeutralSentiment] [Topic]работать[/Topic].

Weak Negative Sentiment

- Первое впечатление - шумно.
Первое впечатление - [WeakNegativeSentiment]шумно[/WeakNegativeSentiment].

Strong Negative Sentiment

- Покрытие чаши жуткое.
[Topic] Покрытие чаши [/Topic] [StrongNegativeSentiment] жуткое [/StrongNegativeSentiment].

Minor Problem

- Пластиковый байонет и хлипкая конструкция.
Пластиковый байонет и [MinorProblem]хлипкая[/MinorProblem] [Topic]конструкция[/Topic].

Major Problem

- Но к сожалению недостатки перечисленные выше испортили представление.
Но к сожалению недостатки перечисленные выше [MajorProblem]испортили[/MajorProblem] [Topic]представление[/Topic].

6.1.1.10 Spanish Sentiment Extraction Examples

Examples of the extraction of positive, neutral, and negative feelings in Spanish.

Strong Positive Sentiment

- Absolutamente adoro este álbum.
Absolutamente [StrongPositiveSentiment] adoro[/StrongPositiveSentiment] este [Topic] álbum[/Topic].

Weak Positive Sentiment

- Me gusta este grupo.
Me [WeakPositiveSentiment] gusta[/WeakPositiveSentiment] este [Topic] grupo[/Topic].

Neutral Sentiment

- Este coche funciona normalmente.
Este [Topic] coche[/Topic] funciona [NeutralSentiment] normalmente[/NeutralSentiment].

Weak Negative Sentiment

- Es una mala tienda.
Es una [WeakNegativeSentiment] mala[/WeakNegativeSentiment] [Topic] tienda[/Topic].

Strong Negative Sentiment

- Odio este televisor.
[StrongNegativeSentiment] Odio[/StrongNegativeSentiment] este [Topic] televisor[/Topic].

Minor Problem

- Tengo problemas con el sonido.

Tengo [MinorProblem]problemas[/MinorProblem] con el [Topic] sonido[/Topic].

Major Problem

- El archivo es corrupto.
El [Topic] archivo[/Topic] es [MajorProblem] corrupto[/MajorProblem].

6.1.2 Emoticon Extraction

Emoticon rules are designed to extract information about someone's feelings about the whole sentence or situation.

Unlike sentiment rules, emoticon rules do not specify a particular topic of the stance.

The rules categorize extracted emoticons or emojis into the following types of stances:

Type of Emoticon	Description
Strong positive emoticon	Extracts emoticons conveying strong positive sentiment, e.g. :-D, :-)), (((:
Weak positive emoticon	Extracts emoticons conveying weak positive sentiment, e.g. :-), :-), :), (:
Neutral emoticon	Extracts emoticons conveying neutral sentiment, e.g., :- , :- , :o, O_O
Weak negative emoticon	Extracts emoticons conveying weak negative sentiment, e.g. :-), :(, :-\
Strong negative emoticon	Extracts emoticons conveying strong negative sentiment, e.g. :-(((, :((, :-(-

6.1.2.1 Arabic Emoticon Extraction Example

Examples of positive and negative emoticons in Arabic.

Strong Positive Emoticon

- كان العرض ممتازا ☺

StrongPositiveEmoticon

☺

Weak Positive Emoticon

- كان العرض جيدا ☺

WeakPositiveEmoticon

☺

Neutral Emoticon

- العرض كان لا بأس به :-|

NeutralEmoticon

:-|

Weak Negative Emoticon

- -_- كان العرض سيئا

WeakNegativeEmoticon

-_-

Strong Negative Emoticon

- (""""": كان العرض فظيحا

StrongNegativeEmoticon

)""""":

6.1.2.2 Chinese (Simplified) Emoticon Extraction Examples

Examples of positive and negative emoticons in Simplified Chinese.

Strong Positive Emoticon

- 这部电影真精彩:D
[Emoticon] 这部电影真精彩 [StrongPositiveEmoticon]:D[/StrongPositiveEmoticon] [/Emoticon]

Weak Positive Emoticon

- 我们的事业蒸蒸日上:)
[Emoticon] 我们的事业蒸蒸日上 [WeakPositiveEmoticon]:)[/WeakPositiveEmoticon] [/Emoticon]

Neutral Emoticon

- 表现差强人意:|
[Emoticon] 表现差强人意 [NeutralEmoticon]:|[/NeutralEmoticon] [/Emoticon]

Weak Negative Emoticon

- 这部手机不好:(
[Emoticon] 这部手机不好 [WeakNegativeEmoticon]:([/WeakNegativeEmoticon] [/Emoticon]

Strong Negative Emoticon

- 我很讨厌他:'-(
[Emoticon] 我很讨厌他 [StrongNegativeEmoticon]:'-([/StrongNegativeEmoticon] [/Emoticon]

6.1.2.3 Chinese (Traditional) Emoticon Extraction Examples

Examples of positive and negative emoticons in Traditional Chinese.

Strong Positive Emoticon

- 這部電影真精彩 :D
[Emoticon] 這部電影真精彩 [StrongPositiveEmoticon]:D[/StrongPositiveEmoticon] [/Emoticon]

Weak Positive Emoticon

- 我們的事業蒸蒸日上 :)
[Emoticon] 我們的事業蒸蒸日上 [WeakPositiveEmoticon]:)[/WeakPositiveEmoticon] [/Emoticon]

Neutral Emoticon

- 表現差強人意 :|
[Emoticon] 表現差強人意 [NeutralEmoticon]:|[/NeutralEmoticon] [/Emoticon]

Weak Negative Emoticon

- 這部手機不好 :(
[Emoticon] 這部手機不好 [WeakNegativeEmoticon]:([/WeakNegativeEmoticon] [/Emoticon]

Strong Negative Emoticon

- 我很討厭他 :'
[Emoticon] 我很討厭他 [StrongNegativeEmoticon]:'([/StrongNegativeEmoticon] [/Emoticon]

6.1.2.4 Dutch Emoticon Extraction Examples

Examples of positive and negative emoticons in Dutch.

Strong Positive Emoticon

- :-D Ik lach hardop
[Emoticon] [StrongPositiveEmoticon] :-D[/StrongPositiveEmoticon] Ik lach hardop[/Emoticon]

Weak Positive Emoticon

- :-) Ik ben blij
[Emoticon] [WeakPositiveEmoticon] :-)[/WeakPositiveEmoticon] Ik ben blij[/Emoticon]

Neutral Emoticon

- :-| Het doet me niks
[Emoticon] [NeutralEmoticon] :-|[/NeutralEmoticon] Het doet me niks[/Emoticon]

Weak Negative Emoticon

- :-(Ik ben verdrietig
[Emoticon] [WeakNegativeEmoticon] :-([/WeakNegativeEmoticon] Ik ben verdrietig[/Emoticon]

Strong Negative Emoticon

- :-((Ik ben heel erg boos
[Emoticon] [StrongNegativeEmoticon] :-(([/StrongNegativeEmoticon] Ik ben heel erg boos[/Emoticon]

6.1.2.5 English Emoticon Extraction Examples

Examples of positive and negative emoticons in English.

Strong Positive Emoticon

- The show was hilarious :-D

[Emoticon] The show was hilarious [StrongPositiveEmoticon]:-D[/StrongPositiveEmoticon]
[/Emoticon]

Weak Positive Emoticon

- Loving my new BlackBerry! :-) No iPhone needed over here.
[Emoticon]Loving my new BlackBerry! [WeakPositiveEmoticon]:-)[/WeakPositiveEmoticon] No
iPhone needed over here. [/Emoticon]

Neutral Emoticon

- The show was not great but it was not terrible either :-| It was ok I guess.
- [Emoticon]The show was not great but it was not terrible either [NeutralEmoticon]:-|[/
NeutralEmoticon] It was ok I guess. [/Emoticon]

Weak Negative Emoticon

- I hate this phone I'm using :-(
[Emoticon] I hate this phone I'm using [WeakNegativeEmoticon]:-([/WeakNegativeEmoticon] [/
Emoticon]

Strong Negative Emoticon

- The Dow Jones fell 200 points :-(((goodbye savings
[Emoticon] The Dow Jones fell 200 points [StrongNegativeEmoticon] :-((([/
StrongNegativeEmoticon] goodbye savings [/Emoticon]

6.1.2.6 French Emoticon Extraction Examples

Examples of positive and negative emoticons in French.

Strong Positive Emoticon

- C'est marrant ce truc! :-D

[Emoticon]C'est marrant ce truc! [StrongPositiveEmoticon]:-D[/StrongPositiveEmoticon] [/Emoticon]

Weak Positive Emoticon

- J'adore mon nouvel iphone! :-)
[Emoticon]J'adore mon nouvel iphone! [WeakPositiveEmoticon]:-)[/WeakPositiveEmoticon] [/Emoticon]

Neutral Emoticon

- Je travaille demain :-|
[Emoticon]Je travaille demain [NeutralEmoticon] :-|[/NeutralEmoticon] [/Emoticon]

Weak Negative Emoticon

- Je n'aime pas ce nouveau GSM :-(
[Emoticon] Je n'aime pas ce nouveau GSM [WeakNegativeEmoticon]:-([/WeakNegativeEmoticon] [/Emoticon]

Strong Negative Emoticon

- Les vacances sont finies :-(((
[Emoticon]Les vacances sont finies [StrongNegativeEmoticon] :-((([/StrongNegativeEmoticon] [/Emoticon]

6.1.2.7 German Emoticon Extraction Examples

Examples of positive and negative emoticons in German.

Strong Positive Emoticon

- Wie krass ihr seid! :-D
[Emoticon] Wie krass ihr seid! [StrongPositiveEmoticon]:-D[/StrongPositiveEmoticon] [/Emoticon]

Weak Positive Emoticon

- Danke für 500 Twitterfollower :-)
[Emoticon]Danke für 500 Twitterfollower [WeakPositiveEmoticon]:-)[/WeakPositiveEmoticon]
[/Emoticon]

Neutral Emoticon

- keine Ahnung was ich essen soll :|
[Emoticon]keine Ahnung was ich essen soll [NeutralEmoticon]:|[/NeutralEmoticon] [/Emoticon]

Weak Negative Emoticon

- Ich verstehe mein Handy nicht!!! :-(
[Emoticon]Ich verstehe mein Handy nicht!!! [WeakNegativeEmoticon]:-([/WeakNegativeEmoticon] [/Emoticon]

Strong Negative Emoticon

- Mir gehts so schlecht :'
[Emoticon]Mir gehts so schlecht [StrongNegativeEmoticon]:'([/StrongNegativeEmoticon] [/Emoticon]

6.1.2.8 Italian Emoticon Extraction Examples

Examples of positive, neutral, and negative emoticons in Italian.

Strong Positive Emoticon

- Uno spettacolo davvero divertente! =D
[Emoticon]Uno spettacolo davvero divertente! [StrongPositiveEmoticon]=D[/StrongPositiveEmoticon] [/Emoticon]

Weak Positive Emoticon

- Sono arrivate le vacanze ☺
[Emoticon] Sono arrivate le vacanze [WeakPositiveEmoticon] ☺ [WeakPositiveEmoticon] [Emoticon]

Neutral Emoticon

- Domani lavoro :|
[Emoticon] [NeutralEmoticon] :| [NeutralEmoticon] [Emoticon]

Weak Negative Emoticon

- Uffa :(
[Emoticon] Uffa [WeakNegativeEmoticon] :([WeakNegativeEmoticon] [Emoticon]

Strong Negative Emoticon

- Che peccato! :(((
[Emoticon] Che peccato! [StrongNegativeEmoticon] :((([StrongNegativeEmoticon] [Emoticon]

6.1.2.9 Portuguese Emoticon Extraction Examples

Examples of positive, neutral, and negative emoticons in Portuguese.

Strong Positive Emoticon

- Estou contentíssima :-D
[Emoticon] Estou contentíssima [StrongPositiveEmoticon] :-D [StrongPositiveEmoticon] [Emoticon]

Weak Positive Emoticon

- Estou feliz :-)
`[Emoticon]Estou feliz [WeakPositiveEmoticon] :-)[/WeakPositiveEmoticon] [/Emoticon]`

Neutral Emoticon

- Estou indiferente :|
`[Emoticon]Estou indiferente [NeutralEmoticon] :|[/NeutralEmoticon] [/Emoticon]`

Weak Negative Emoticon

- Estou triste :-(
`[Emoticon]Estou triste [WeakNegativeEmoticon] :-([/WeakNegativeEmoticon] [/Emoticon]`

Strong Negative Emoticon

- Estou tristíssima :(((
`[Emoticon]Estou tristíssima [StrongNegativeEmoticon] :((([/StrongNegativeEmoticon] [/Emoticon]`

6.1.2.10 Russian Emoticon Extraction Examples

Examples of positive and negative emoticons in Russian.

Strong Positive Emoticon

- Шоу было очень смешное :-)))
`[Emoticon] Шоу было очень смешное [StrongPositiveEmoticon] :-)))[/StrongPositiveEmoticon] [/Emoticon]`

Weak Positive Emoticon

- А он красавчик! :-)

[Emoticon] А он красавчик! [WeakPositiveEmoticon] :-) [/WeakPositiveEmoticon] [/Emoticon]

Neutral Emoticon

- Фильм был нормальный :- |
[Emoticon] Фильм был нормальный [NeutralEmoticon] :- | [/NeutralEmoticon] [/Emoticon]

Weak Negative Emoticon

- Терпеть не могу пауков :-(
[Emoticon] Терпеть не могу пауков [WeakNegativeEmoticon] :-([/WeakNegativeEmoticon] [/Emoticon]

Strong Negative Emoticon

- Получил два балла на экзамене :-(
[Emoticon] Получил два балла на экзамене [StrongNegativeEmoticon] :-(([/StrongNegativeEmoticon] [/Emoticon]

6.1.2.11 Spanish Emoticon Extraction Examples

Examples of positive and negative emoticons in Spanish.

Strong Positive Emoticon

- Nos vemos mas tarde! :-D
[Emoticon] Nos vemos mas tarde! [StrongPositiveEmoticon] :-D [/StrongPositiveEmoticon] [/Emoticon]

Weak Positive Emoticon

- Puedo venir a visitarte! :-)
[Emoticon] Puedo venir a visitarte! [WeakPositiveEmoticon] :-) [/WeakPositiveEmoticon] [/Emoticon]

Neutral Emoticon

- Trabajo hoy :-|
[Emoticon] Trabajo hoy [NeutralEmoticon] :-|[NeutralEmoticon] [/Emoticon]

Weak Negative Emoticon

- No me gusta este celular :-(
[Emoticon] No me gusta este celular [WeakNegativeEmoticon] :-([WeakNegativeEmoticon] [/Emoticon]

Strong Negative Emoticon

- Se acaban las vacaciones :-(((
[Emoticon] Se acaban las vacaciones [StrongNegativeEmoticon] :-((([StrongNegativeEmoticon] [/Emoticon]

6.1.3 Request Extraction

Request rules are designed to extract information about a customer's wish for a change or enhancement. The rules extract patterns that express customer requests to be contacted or for new or added functionality in an item such as a product, company, service, or person.

Requests are categorized into the following stances:

Type of Request Stance	Description
General request	A request for an enhancement or for something new, such as "please add", "please create", or "would like"
Contact request	A request for direct or immediate contact, such as "please contact me" or "call me"

An extracted request contains the following:

- One request stance (`ContactRequest` or `GeneralRequest`)
- Optionally, one topic—what the the request is about. The topic answers the question "What does the customer want?"
- Optionally, contact information—for phone or fax numbers, addresses, e-mail addresses, web site addresses

6.1.3.1 Arabic Request Extraction Examples

Examples of general requests and contact requests in Arabic.

General Requests

- ممكن تزودوا زر؟

General Request	ممکن
Topic	تزودوا زر

Contact Requests

- ممكن الكتالوج؟

Contact Request	ممکن
Topic	الكتالوج

6.1.3.2 Chinese (Simplified) Request Extraction Examples

Examples of general requests and contact requests in Simplified Chinese

General Requests

- 请尽快完成任务。
[GeneralRequest] 请 [/GeneralRequest] 尽快完成 [Topic] 任务 [/Topic]。

Contact Requests

- 请联系我们。
请 [ContactRequest] 联系 [/ContactRequest] 我们。
- 我们的电话是 (025)83498215
我们的 [ContactRequest] 电话 [/ContactRequest] 是 [ContactInfo](025)83498215 [ContactInfo]。

6.1.3.3 Chinese (Traditional) Request Extraction Examples

Examples of general requests and contact requests in Traditional Chinese

General Requests

- 請儘快完成任務。
[GeneralRequest]請[/GeneralRequest]儘快完成 [Topic] 任務[/Topic]。

Contact Requests

- 請聯繫我們。
請 [ContactRequest] 聯繫 [/ContactRequest] 我們。
- 我們的電話是 (025)83498215。
我們的 [ContactRequest] 電話 [/ContactRequest] 是 [ContactInfo](025)83498215 [ContactInfo]。

6.1.3.4 English Request Extraction Examples

Examples of general requests and contact requests in English.

General Requests

- Improve the software UI.
[GeneralRequest]Improve[/GeneralRequest] the [Topic] software UI [Topic].
- An additional switch would be great to have on this vacuum cleaner.
An [Topic] additional switch[/Topic] [GeneralRequest]would be great[/GeneralRequest] to have on this vacuum cleaner.

Contact Requests

- I would like to be contacted by your customer support service.
I [ContactRequest]would like to be contacted[/ContactRequest] by your [Topic] customer support service [Topic].
- I would like to receive the January catalog.
I [ContactRequest]would like to receive[/ContactRequest] the [Topic] January catalog [Topic].

6.1.3.5 French Request Extraction Examples

Examples of general requests and contact requests in French

General Requests

- Le conso aimerait une extension de garantie.
Le conso [GeneralRequest]aimerait[/GeneralRequest] [Topic] une extension de garantie[/Topic].
- J'aurais aimé trouver plus de fonctions.
J'[GeneralRequest]aurais aimé[/GeneralRequest] [Topic]trouver plus de fonctions[/Topic].

Contact Requests

- Il souhaitait des informations sur les nouveaux produits.
Il [ContactRequest]souhaitait[/ContactRequest] [Topic]des informations sur les nouveaux produits[/Topic].
- Il demande des renseignements sur la garantie constructeur.
Il [ContactRequest]demande[/ContactRequest] [Topic]des renseignements sur la garantie constructeur[/Topic].

6.1.3.6 German Request Extraction Examples

Examples of general requests and contact requests in German

General Requests

- Lizenzmodel deutlich vereinfachen!
[Topic]Lizenzmodel[/Topic]deutlich [GeneralRequest]vereinfachen[/GeneralRequest]!

Contact Requests

- Rufen Sie mich bitte sofort an.
[ContactRequest]Rufen Sie[/ContactRequest] mich bitte sofort an.

6.1.3.7 Italian Request Extraction Examples

Examples of general requests and contact requests in Italian

General Requests

- Mi sarebbe piaciuto avere la funzione di crossover elettronico.
Mi [GeneralRequest] sarebbe piaciuto[/GeneralRequest] [Topic] avere la funzione di crossover elettronico [/Topic].
- Sarebbe ora che faceste una web app
[GeneralRequest] Sarebbe ora [/GeneralRequest] che [Topic] faceste una web app [/Topic]

Contact Requests

- Vorrei essere contattato dal vostro servizio assistenza.
[ContactRequest] Vorrei essere contattato [/ContactRequest] dal [Topic] vostro servizio assistenza [/Topic].
- Potete mandarmi il catalogo invernale?
[ContactRequest] Potete mandarmi [/ContactRequest] [Topic] il catalogo invernale [/Topic]?

6.1.3.8 Portuguese Request Extraction Examples

Examples of general requests and contact requests in Portuguese

General Requests

- Gostaria de poder trocar a cor.
[GeneralRequest] Gostaria [/GeneralRequest] de [Topic] poder trocar a cor [/Topic].

Contact Requests

- Por favor contacte-me pelo telefone: 508 653-4455.
Por favor [ContactRequest] contacte-me [/ContactRequest] pelo [ContactInfo] telefone: 508 653-4455 [/ContactInfo].

6.1.3.9 Russian Request Extraction Examples

Examples of general requests and contact requests in Russian

General Requests

- Добавьте кнопку "Отмена"
[GeneralRequest]Добавьте[/GeneralRequest] [Topic]кнопку[/Topic] "Отмена"
- Пожалуйста перезагрузите сервер
Пожалуйста [GeneralRequest]перезагрузите[/GeneralRequest] [Topic]сервер[/Topic]

Contact Requests

- Пожалуйста, свяжитесь со мной по телефону
Пожалуйста, [ContactRequest]свяжитесь[/ContactRequest] со мной по телефону
- Отправьте ваш каталог пожалуйста
[ContactRequest]Отправьте[/ContactRequest] [Topic]ваш каталог[/Topic] пожалуйста

6.1.3.10 Spanish Request Extraction Examples

Examples of general requests and contact requests in Spanish

General Requests

- Podría hacer un otro color en vez de blanco?
[GeneralRequest]Podría[/GeneralRequest] [Topic]hacer un otro color en vez de blanco[/Topic]?

Contact Requests

- Quiero contactarme contigo.
[ContactRequest]Quiero[/ContactRequest] [Topic]contactarme contigo[/Topic].

6.1.4 Profanity Extraction

The aim of the Profanity dictionary is to define a set of pejorative vocabulary.

Profanities are categorized into two classes:

Type of Dictionary	Description
Ambiguous profanity	Extracts words and phrases that are pejorative only in certain contexts
Unambiguous profanity	Extracts words and phrases that are always pejorative, regardless of the context

This functionality is supported in the following languages:

- Chinese (Simplified)
- Chinese (Traditional)
- Dutch
- English
- French
- German
- Italian
- Portuguese
- Russian
- Spanish

6.2 Public Sector Fact Extraction

The public sector content includes a set of rules that you can use to extract public-sector-specific information when processing and analyzing text.

It is included in, and supports only, the English language module.

When used in conjunction with the standard English language module, the public sector content enables you to extract public-sector-specific entities in addition to the default entity types. It includes rule sets that allow for the extraction of additional public-sector-related information.

The public sector content provides these extraction abilities:

- Default entity types, as documented in the "Entity Extraction" section of this guide
- Public-sector-specific facts and entities
- Common mentions

Rule Set Description

- Action
Extracts information about action and travel events

- Military Units
Extracts information about military units such as teams, wings, and squadrons

- Organizational Information
Extracts information about organizations

- Person-Alias
Extracts information about a person's possible aliases

- Person-Appearance
Extracts information about a person's appearance

- Person-Attributes
Extracts information about a person's non-appearance attributes

- Person-Relationships
Extracts information about a person's relationships

- Spatial References
Extracts relative spatial references, such as distances, cardinal directions, or locations

For more details about writing your own extraction rules, refer to the *SAP Data Services Text Data Processing Extraction Customization Guide*.

For a full list of entity and subentity names, see [Entity Type Names and Subentity Type Names For Fact Extraction \[page 378\]](#) and note those whose name begins with "PublicSector".

Related Information

[Action Event Extraction \[page 353\]](#)

[Travel Event Extraction \[page 365\]](#)

[Military Unit Extraction \[page 355\]](#)

[Organizational Information Extraction \[page 356\]](#)

[Person: Alias Extraction \[page 357\]](#)

[Person: Appearance Extraction \[page 359\]](#)

[Person: Attribute Extraction \[page 361\]](#)

[Person: Relationship Extraction \[page 362\]](#)

[Spatial Reference Extraction \[page 364\]](#)

6.2.1 Common Entity Extraction

Common noun mentions refers to the use of common nouns to designate entities such as organizations, persons, or facilities which would normally also be referred to by proper nouns. They are defined as noun phrases headed by an appropriate noun. Both singular and plural forms are matched. Proper nouns and modifiers are also included.

Common mentions content includes a set of entity types and rules that address requirements for extracting common mentions. You can use this content to extract these specific types of information when processing and analyzing text:

Rule Set Description

- Common Mentions
Extracts information about common noun mentions

Note

The common mentions content is included in and supports the English language module only.

For details about using common mentions content to enhance extraction rules, refer to the *SAP Data Services Text Data Processing Extraction Customization Guide*.

Type and Possible Subtype	Definition	Examples
COMMON_ADDRESS1 and COMMON_ADDRESS2	Common nouns for addresses, mirroring their named counterparts even if there is not any real difference in content between them	fictitious address current address
COMMON_CONTINENT	Common nouns for the entirety of any continent	major continents
COMMON_COUNTRY	Common nouns for the entirety of any country, including disputed territories	beloved motherland major countries Native American reservation smaller nations
COMMON_FACILITY@AIRPORTS	Common nouns for air transportation terminals	commercial airport busy air field public heliport

Type and Possible Subtype	Definition	Examples
COMMON_FACILITY@BUILDGROUNDS	Common nouns for architectural and civil engineering structures, and outdoor spaces	public library famous national archives national park training camp train station naval port
COMMON_FACILITY@PATH	Common nouns for man-made structures that allow fluids, energy, persons, animals, or vehicles to pass from one location to another	deserted street narrow canal heavily defended bridge
COMMON_FACILITY@PLANT	Common nouns for man-made structures used for industrial purposes	oil refinery copper smelter thermal power station steel foundry
COMMON_FACILITY@SUBAREA	Common nouns for portions of man-made facilities able to contain people, animals, or objects	small atrium cold cellar new kitchen top-floor apartment
COMMON_GEO_AREA@DOMESTIC	Common nouns for non-political geographical areas capturing a significant land mass that do not cross national borders	remote region open frontier area
COMMON_GEO_AREA@INTL	Common nouns for non-political geographical areas capturing a significant land mass that do cross international borders	overseas
COMMON_GEO_FEATURE@BOUNDARY	Common nouns for non-political geographical locations such as a border	northern border unaccessible frontiers

Type and Possible Subtype	Definition	Examples
COMMON_GEO_FEATURE@CELESTIAL	Common nouns for non-political geographical locations outside of Earth, such as planets	largest planet night sky
COMMON_GEO_FEATURE@LAND	Common nouns for non-political geographical locations: bodies of land, such as deserts, mountains, volcanoes, plateaus, plains, peninsulas, canyons, beaches, etc.	mountain range French seaside
COMMON_GEO_FEATURE@WATER	Common nouns for non-political geographical locations: bodies of water	saltwater lake flooding rivers
COMMON_LOCALITY	Common nouns for cities	border town densely populated cities
COMMON_ORGANIZATION@COMMERCIAL	Common nouns for companies	small robotics company pesticides manufacturers world's fourth-biggest airline
COMMON_ORGANIZATION@EDUCATIONAL	Common nouns for institutions focused on education	private university public colleges
COMMON_ORGANIZATION@ENTERTAINMENT	Common nouns for institutions focused on entertainment	contemporary circus theater company
COMMON_ORGANIZATION@GOVERNMENT	Common nouns for institutions related to government, politics, or the state	Obama government battalion justice system
COMMON_ORGANIZATION@MEDIA	Common nouns for institutions related to the media	news service television station
COMMON_ORGANIZATION@MEDICALSCIENCE	Institutions related to medicine or research	health group teaching hospital
COMMON_ORGANIZATION@RELIGIOUS	Common nouns for institutions related to religion	Catholic church powerful archdiocese

Type and Possible Subtype	Definition	Examples
COMMON_ORGANIZATION@SPORTS	Common nouns for institutions related to sports	major league sport team
COMMON_ORGANIZATION@OTHER	Common nouns for organizations that do not fit into a more specific subtype	Palestinian and Lebanese organizations largest opposition party
COMMON_PERSON	Common nouns for persons	ceremony ministers injured members submarine crew 58-year-old man math teacher
COMMON_PRECURSOR@CHEMICAL	Common names of chemical substances that are used in the manufacturing of weapons	precursor chemical material
COMMON_PRECURSOR@NUCLEAR	Common names of nuclear and radiological substances used in the manufacturing of weapons	precursor nuclear material
COMMON_REGION@MAJOR	Common nouns of administrative units above the city and below the country: states and provinces	historical provinces home state
COMMON_REGION@MINOR	Common nouns of administrative units above the city and below the country: counties and districts	millionaire counties development district
COMMON_VEHICLE@AIR	Common names for air vehicles	dirigible black helicopter jetliner
COMMON_VEHICLE@LAND	Common names for land vehicles	car motorbike

Type and Possible Subtype	Definition	Examples
COMMON_VEHICLE@WATER	Common names for water vehicles	cruiseline boat Russian submarine
COMMON_VEHICLE@SUBAREA	Common names for portions of vehicles in which humans can fit	cockpit engine room
COMMON_VEHICLE@OTHER	Common names for vehicles that do not fit into a more specific subtype	chopper
COMMON_WEAPON@BIOLOGICAL	Common names for biological weapons	attractive biological threat agent
COMMON_WEAPON@CHEMICAL	Common names for chemical weapons	nerve agent
COMMON_WEAPON@EXPLODING	Common names for conventional weapons designed to explode	rocket-propelled grenades
COMMON_WEAPON@NUCLEAR	Common names for nuclear weapons	radioactive materials
COMMON_WEAPON@PROJECTILE	Common names for conventional weapons designed to be projected at great speed	bullets
COMMON_WEAPON@SHARP	Common names for conventional weapons designed to cut	machetes
COMMON_WEAPON@SHOOTING	Common names for conventional weapons designed to send projectile objects at great speed	rifle
COMMON_WEAPON@OTHER	Common names for conventional weapons that do not fit into a more specific subtype	weapons

6.2.2 Action Event Extraction

Extracts information related to events involving persons or organizations and their movements, creations, and transfers.

Separate rules have been created to extract actions described not only in the active but also in the passive voice, where applicable. The table below shows the types of events extracted as Action events.

Types of Events Extracted as Action Events

Buy	Destroy	Hire	Make	Train
Capture	Drive	Indict	Participate	Transport
Command	Execute	Injure	Receive_Pay	
Communicate	Finance	Kill	Survey	

Information types

The specific types of information extracted by these rules are the following:

Type	Definition
Agent	First of two arguments taking part in an action event or a travel event
Artifact	Second of two arguments taking part in an action event (inanimate)
Date/Time/Place	Date, time, or place expression occurring in same sentence as an action or travel event
FromPlace	Origin of trip in a travel event
Organization	Organization taking part in an action event
Patient	Second of two arguments taking part in an action event (animate)
Payment	Payment given to Recipient
Recipient	Person or organization receiving payment
ToPlace	Destination in a travel event
Travel	Travel movement in a travel event
Vehicle	Vehicle taking part in an action event

These information types are the entity subtypes are extracted. The event types contribute to the naming of the events, such as `Action_Drive_Active`, `Action_Train_With`, or `Action_Receive_Pay_Passive`.

Examples

- Original Fathi Subuh was arrested by the Palestinian Authority's Preventative Security Service (PSS) on July 2, 1997.
- Extracted **[Action_Capture_Passive]** **[Patient]**Fathi Subuh**[/Patient]** was arrested by the **[Agent]**Palestinian Authority's Preventative Security Service**[/Agent]** (PSS) on **[Date]**July 2, 1997**[/Date]**. **[/Action_Capture_Passive]**
- Original Pakistan President Gen. Pervez Musharraf coordinated the Pakistani army efforts last April.
- Extracted **[Action_Command_Active]** **[Agent]**Pakistan President Gen. Pervez Musharraf**[/Agent]** coordinated the **[Organization]**Pakistani army**[/Organization]** efforts last **[Date]**April**[/Date]**. **[/Action_Command_Active]**
- Original The Third Regiment, Texas Volunteer Infantry (United States Volunteers), was commanded by Colonel R. P. Smyth.
- Extracted **[Action_Command_Passive]**The **[Organization]**Third Regiment, Texas Volunteer Infantry**[/Organization]** (United States Volunteers), was commanded by **[Agent]**Colonel R. P. Smyth**[/Agent]**. **[/Action_Command_Passive]**
- Original In April, Iraqi insurgents attacked 5 US Army convoys.
- Extracted **[Action_Destroy_Active]**In **[Date]**April**[/Date]**, **[Agent]**Iraqi insurgents**[/Agent]** attacked **[Artifact]**5 US Army convoys**[/Artifact]**. **[/Action_Destroy_Active]**
- Original Jane drives a green Subaru to work.
- Extracted **[Action_Drive_Active]** **[Agent]**Jane**[/Agent]** drives a **[Vehicle]**green Subaru**[/Vehicle]** to work. **[/Action_Drive_Active]**
- Original Daniel Pearl was executed by his captors in Karachi.
- Extracted **[Action_Execute_Passive]** **[Patient]**Daniel Pearl**[/Patient]** was executed by his **[Agent]**captors**[/Agent]** in **[Place]**Karachi**[/Place]**. **[/Action_Execute_Passive]**
- Original Some Islamic organizations might be financed by wealthy Saudis.
- Extracted **[Action_Finance_Passive]** **[Patient]**Some Islamic organizations**[/Patient]** might be financed by wealthy **[Agent]**Saudis**[/Agent]**. **[/Action_Finance_Passive]**
- Original The local soccer federation hired John Brown as acting director.

Extracted **[Action_Hire_Active]** **[Agent]**The local soccer federation**[/Agent]** hired **[Patient]**John Brown**[/Patient]** as acting director.**[/Action_Hire_Active]**

Original In 2000, Milosevic was indicted by the United Nations.

Extracted **[Action_Indict_Passive]**In **[Date]**2000**[/Date]**, **[Patient]**Milosevic**[/Patient]** was indicted by the **[Agent]**United Nations**[/Agent]**.**[/Action_Indict_Passive]**

Original Jamal Ahmed Al-Fadl received \$10,000 for his time and effort and did not take a further role in the uranium acquisition.

Extracted **[Action_Receive_Pay_Active]** **[Recipient]**Jamal Ahmed Al-Fadl**[/Recipient]** received **[Payment]**\$10,000**[/Payment]** for his time and effort and did not take a further role in the uranium acquisition.**[/Action_Receive_Pay_Active]**

Original The Montreal Baseball School trained John before he played in New York.

Extracted **[Action_Train_Active]** **[Agent]**The Montreal Baseball School**[/Agent]** trained **[Patient]**John**[/Patient]** before he played in **[Place]**New York**[/Place]**.**[/Action_Train_Active]**

Original Over the past 3 years, presumed terrorists were transported by the US Government to an undisclosed location.

Extracted **[Action_Transport_Passive]**Over the past **[Time]**3 years**[/Time]**, **[Patient]**presumed terrorists**[/Patient]** were transported by the **[Agent]**US Government**[/Agent]** to an undisclosed location.**[/Action_Transport_Passive]**

6.2.3 Military Unit Extraction

The `MilitaryUnit` rule is designed to extract expressions that refer to military units, including long and coordinated expressions.

Examples

Original Paul serves in Company A, 1st Battalion, 22nd Infantry Regiment, 1st Brigade Combat Team, 4th Infantry Division.

Extracted Paul serves in **[MilitaryUnit]**Company A, 1st Battalion, 22nd Infantry Regiment, 1st Brigade Combat Team, 4th Infantry Division**[/MilitaryUnit]**.

Original Soldiers from Company C, 3rd Battalion, 67th Armor Regiment, 4th Brigade Combat Team, 101st Airborne Division, detained four kidnappers in Baghdad July 31.

Extracted Soldiers from **[MilitaryUnit]**Company C, 3rd Battalion, 67th Armor Regiment, 4th Brigade Combat Team, 101st Airborne Division~~[/MilitaryUnit]~~, detained four kidnappers in Baghdad July 31.

6.2.4 Organizational Information Extraction

Extracts core information about companies and organizations.

Information Types

The specific types of information extracted by these rules include the organization name and at least one of the following:

Type	Definition
Address	Street address of the organization
Email	E-mail address
Fax	Fax number, with extension
Location	Location of the organization
Organization	Organization about which the information is collected
Person	Individual to contact, including title and department where provided
Phone	Phone number, with extension
Url	Web Address

Examples

Original Patti J. McAtee, Director of CalEnergy, 402-341-4500.

Extracted **[OrganizationInfo]****[Person]**Patti J. McAtee~~[/Person]~~, Director of **[Organization]**CalEnergy~~[/Organization]~~, **[Phone]**402-341-4500~~[/Phone]~~ **[OrganizationInfo]**.

Original Marriott International press releases are available through Company News On-Call by fax, 800-758-5804, ext. 532963.

Extracted **[OrganizationInfo][Organization]**Marriott International**[/Organization]** press releases are available through Company News On-Call by fax, **[Fax]**800-758-5804, ext. 532963**[/Fax][/OrganizationInfo]**.

Original Inxight Software Inc., 500 Macara Avenue, Sunnyvale, CA 94085, U.S.A., Email: info@inxight.com

Extracted **[OrganizationInfo][Organization]**Inxight Software Inc.**[/Organization]**, **[Address]**500 Macara Avenue, Sunnyvale, CA 94085, U.S.A.**[/Address]**, Email: **[Email]**info@inxight.com**[/Email][/OrganizationInfo]**

Original New York-based American Express Co. is best known for its credit card.

Extracted **[OrganizationInfo][Location]**New York-based**[/Location]** **[Organization]**American Express Co.**[/Organization][/OrganizationInfo]** is best known for its credit card.

Original Michelin Tyre is a unit of France's Michelin S.A.

Extracted Michelin Tyre is a unit of **[OrganizationInfo][Nationality]**France**[/Nationality]**'s **[Organization]**Michelin S.A.**[/Organization][/OrganizationInfo]**

Original Microsoft (Nasdaq:MSFT) announced this quarter's earnings.

Extracted **[OrganizationInfo][Organization]**Microsoft**[/Organization]** (**[Ticker]**Nasdaq:MSFT**[/Ticker]**)**[/OrganizationInfo]** announced this quarter's earnings.

Original Kodak shares were up 4.5 percent to \$45.87.

Extracted **[OrganizationInfo][Organization]**Kodak**[/Organization]** shares were up 4.5 percent to **[Price]**\$45.87**[/Price][/OrganizationInfo]**.

6.2.5 Person: Alias Extraction

Extracts information about individuals and the possible alternate names and aliases they might use.

Information Types

The specific types of information extracted by these rules are the following:

Type	Description
Alias	An alias used by a person
Person	The name of a person

Examples

- Original Sam Sneed, a.k.a. William Smith, is one of the group.
- Extracted **[PersonAlias_alias_Proper][Person]Sam Sneed[/Person], a.k.a. [Alias]William Smith[/Alias][/PersonAlias_alias_Proper], is one of the group.**
- Original Soldiers spotted the groups led by another Abu Sayyaf leader, Mr. Umbra Jumdail, or Dr. Abu Pula, as he's known.
- Extracted Soldiers spotted the groups led by another Abu Sayyaf leader, **[PersonAlias_or][Person]Mr. Umbra Jumdail[/Person], or [Alias]Dr. Abu Pula[/Alias],[/PersonAlias_or] as he's known.**
- Original Nasr Fahmi Nasr Hassanein, known as Mohamed Salah, was also involved.
- Extracted **[PersonAlias_AlsoKnownAs][Person]Nasr Fahmi Nasr Hassanein[/Person], known as [Alias]Mohamed Salah[/Alias], was also involved.[/PersonAlias_AlsoKnownAs]**
- Original ABU SUHAYB AL-AMRIKI: A 25-year-old U.S. citizen who is known as Adam Pearlman.
- Extracted **[PersonAlias_AlsoKnownAs_NPWho][Person]ABU SUHAYB AL-AMRIKI[/Person]: A 25-year-old U.S. citizen who is known as [Alias]Adam Pearlman[/Alias].[/PersonAlias_AlsoKnownAs_NPWho]**
- Original Ahmed Khalhan Ghailiani, also known as "Ahmed the Tanzanian," "Foopie," and "Fupi."
- Extracted **[PersonAlias_AlsoKnownAs_Quote][Person]Ahmed Khalhan Ghailiani[/Person], also known as "[Alias]Ahmed the Tanzanian[/Alias]," "[Alias]Foopie[/Alias]," and "[Alias]Fupi[/Alias]."[/PersonAlias_AlsoKnownAs_Quote]**
- Original Subhi Abdel-Aziz El-Gohari Abu Sittah, who also goes by the name Abu Hafs El-Masri.
- Extracted **[PersonAlias_UsingTheName_Who][Person]Subhi Abdel-Aziz El-Gohari Abu Sittah[/Person], who also goes by the name [Alias]Abu Hafs El-Masri[/Alias].[/PersonAlias_UsingTheName_Who]**

Original ABU SUHAYB AL-AMRIKI: A 25-year-old U.S. citizen who also goes by the names Adam Pearlman and Adam Gadahn.

Extracted **[PersonAlias_UsingTheNames_NPWho]** **[Person]**ABU SUHAYB AL-AMRIKI**[/Person]**: A 25-year-old U.S. citizen who also goes by the names **[Alias]**Adam Pearlman**[/Alias]** and **[Alias]**Adam Gadahn**[/Alias]**.**[/PersonAlias_UsingTheNames_NPWho]**

Original MICHAEL KAIGHN is using the aliases "D.S." and "Patrick Grogan".

Extracted **[PersonAlias_UsingTheNames_Quote]** **[Person]**MICHAEL KAIGHN**[/Person]** is using the aliases "**[Alias]**D.S.**[/Alias]**" and "**[Alias]**Patrick Grogan**[/Alias]**".**[/PersonAlias_UsingTheNames_Quote]**

Original Also known as Robert, John is the leader of the organization.

Extracted **[PersonAlias_AlsoKnownAsPRE]**Also known as **[Alias]**Robert**[/Alias]**, **[Person]**John**[/Person]** is the leader of the organization.**[/PersonAlias_AlsoKnownAsPRE]**

Original Using the alias "Jafar the Pilot", John is the leader of the organization.

Extracted **[PersonAlias_AlsoKnownAsPRE_Quote]**Using the alias "**[Alias]**Jafar the Pilot**[/Alias]**", **[Person]**John**[/Person]** is the leader of the organization.**[/PersonAlias_AlsoKnownAsPRE_Quote]**

6.2.6 Person: Appearance Extraction

Extracts appearance-related attributes of a person, such as height and eye color, or style of dress.

Information Types

The specific types of information extracted by these rules are the following:

Type	Definition
Age	Person's age
Dress	Person's attire
Eyes	Person's eye color
Hair	Person's hair color and possibly shape

Type	Definition
Height	Person's height (in numbers)
Person	Name of person
Weight	Person's weight (in numbers)

Examples

Original Mary is a 32-year-old registered nurse.

Extracted **[PerApp_Age] [Person]Mary[/Person]** is a **[Age]32-year-old[/Age]** registered nurse. **[/PerApp_Age]**

Original Hamal wears t-shirt and jeans.

Extracted **[PerApp_Dress] [Person]Hamal[/Person]** wears **[Dress]t-shirt and jeans[/Dress]**. **[/PerApp_Dress]**

Original Ahmad has piercing brown eyes.

Extracted **[PerApp_Eyes] [Person]Ahmad[/Person]** has piercing **[Eyes]brown[/Eyes]** eyes. **[/PerApp_Eyes]**

Original JOHN HAS MEDIUM LENGTH BLACK HAIR.

Extracted **[PerApp_Hair] [Person]JOHN[/Person]** HAS **[Hair]MEDIUM LENGTH BLACK[/Hair]** HAIR. **[/PerApp_Hair]**

Original Muhammed is 35 years old, approximately 186 cm tall, weighs 70 kg, has a thin build.

Extracted **[PerApp_Height] [Person]Muhammed[/Person]** is 35 years old, approximately **[Height]186 cm[/Height]** tall **[/PerApp_Height]**, weighs 70 kg, has a thin build.

6.2.7 Person: Attribute Extraction

Extracts non-appearance-related attributes of a person.

Information Types

The specific types of information extracted by these rules are the following:

Type	Description
Artifact	An artifact in person's possession
Nationality	Person's nationality
Occupation	Person's occupation and/or skill
Organization	Person's affiliation
Person	Name of person
Phone	Person's phone number
Residence	Person's address (specific or non-specific)
Vehicle	Person's vehicle

Examples

Original John lived in Montreal.

Extracted **[PerAtt_Location] [Person]John[/Person] lived in [Residence]Montreal[/Residence].[/PerAtt_Location]**

Original Saad Khayyat, who has lived in New Zealand for more than seven years, said the war was about the Middle East being "remade" to suit American interests.

Extracted **[PerAtt_Location_verbSn] [Person]Saad Khayyat[/Person], who has lived in [Residence]New Zealand[/Residence] for more than seven years, said the war was about the Middle East being "remade" to suit American interests. [/PerAtt_Location_verbSn]**

Original Roderick Liddell (telephone: (0)3 88 41 24 92

Extracted **[PerAtt_Phone] [Person]Roderick Liddell[/Person] ([Phone]telephone: (0)3 88 41 24 92[/Phone]) [/PerAtt_Phone]**

Original John drives a red Subaru.

Extracted **[PerAtt_Vehicle][Person]**John**[/Person]** drives a **[Vehicle]**red Subaru**[/Vehicle]**.**[/PerAtt_Vehicle]**

Original JENNIFER CASOLO, A U.S. CITIZEN WAS ARRESTED BY THE SECURITY CORPS.

Extracted **[PerAtt_Nationality][Person]**JENNIFER CASOLO**[/Person]**, A **[Nationality]**U.S.**[/Nationality]** CITIZEN**[/PerAtt_Nationality]** WAS ARRESTED BY THE SECURITY CORPS.

Original Rep. James Walsh, a senior member of the Appropriations Committee.

Extracted **[PerAtt_Affiliation][Person]**Rep. James Walsh**[/Person]**, a senior member of the **[Organization]**Appropriations Committee**[/Organization]****[/PerAtt_Affiliation]**.

Original No water, no electricity, no work, no medicine, said Ali Noor, a retired chemical engineer.

Extracted No water, no electricity, no work, no medicine, said **[PerAtt_Occupation][Person]**Ali Noor**[/Person]**, a **[Occupation]**retired chemical engineer**[/Occupation]****[/PerAtt_Occupation]**.

Original John has a bike.

Extracted **[PerAtt_Possession][Person]**John**[/Person]** has a **[Artifact]**bike**[/Artifact]**.**[/PerAtt_Possession]**

6.2.8 Person: Relationship Extraction

Extracts relationships between two people, whether familial, social, or work-related.

Information Types

The specific types of information extracted by these rules are the following:

Type	Definition
Associate	Name of person's associate
Child	Name of person's child
Parent	Name of person's parent
Relative	Name of person's relative

Type	Definition
Sibling	Name of person's sibling
Spouse	Name of person's spouse

Examples

Original Surjit Kaur's son Prabhjot Singh.

Extracted **[PerRel_ParentChild][Parent]Surjit Kaur[/Parent]'s son [Child]Prabhjot Singh[/Child][PerRel_ParentChild]**.

Original Rana, the daughter of Fateha Gazi.

Extracted **[PerRel_ParentChild][Child]Rana[/Child], the daughter of [Parent]Fateha Gazi[/Parent].[/PerRel_ParentChild]**

Original Paul and John Singh are Surjit Kaur's sons.

Extracted **[PerRel_ParentChild][Child]Paul[/Child] and [Child]John Singh[/Child] are [Parent]Surjit Kaur[/Parent]'s sons[/PerRel_ParentChild]**.

Original Mohamed Kubwa and his half-brother Amina.

Extracted **[PerRel_Sibling][Sibling]Mohamed Kubwa[/Sibling] and his half-brother [Sibling]Amina[/Sibling][PerRel_Sibling]**.

Original John enlisted his brother Joe.

Extracted **[PerRel_Sibling][Sibling]John[/Sibling] enlisted his brother [Sibling]Joe[/Sibling].[/PerRel_Sibling]**

Original Patricia, the wife of John Manningham.

Extracted **[PerRel_Spouse][Spouse]Patricia[/Spouse], the wife of [Spouse]John Manningham[/Spouse].[/PerRel_Spouse]**

Original John asked his wife Donna to help him.

Extracted **[PerRel_Spouse][Spouse]John[/Spouse] asked his wife [Spouse]Donna[/Spouse] to help him.[PerRel_Spouse]**

Original Bin Laden's brother-in-law, Mohammed Jamal Khalifa.

Extracted **[PerRel_Relative][Relative]Bin Laden[/Relative]'s brother-in-law, [Relative]Mohammed Jamal Khalifa[/Relative].[/PerRel_Relative]**

Original John talked to his uncle Charlie.

Extracted **[PerRel_Relative] [Agent] [Relative] John[/Relative] [/Agent]** talked to his
uncle **[Relative] Charlie[/Relative] . [PerRel_Relative]**

Original A curious story was told by Swaran Singh and his friend Sukhdev Singh.

Extracted **[PerRel_Associate]**A curious story was told by **[Associate] Swaran Singh[/Associate]**
[Associate] and his friend **[Associate] Sukhdev Singh[/Associate]** . [PerRel_Associate]

Original John and Fred are roommates.

Extracted **[PerRel_Associate] [Associate] John[/Associate]** and **[Associate] Fred[/Associate]**
[Associate] are roommates. [PerRel_Associate]

6.2.9 Spatial Reference Extraction

Extracts exact or vague spatial references.

The references can be either exact based on distance, direction, and place (30 KM South of Kandahar), or vague based on preposition and place (in Kandahar, near Kandahar).

Types of Spatial Reference Rules

SpatialReference_Exact These rules extract exact relative spatial references (10 km north of Berlin)

SpatialReference_Vague These rules extract vague relative spatial references (near Berlin)

Information Types

The specific types of information extracted by these rules are the following:

Type	Definition
Direction	Keyword indicating a direction (north, near, not too far from, and so on)
Distance	A distance in km, miles
Place	Name of a place used in the relative spatial reference

Examples

Original 15 km north of Kandahar

Extracted **[SpatialReference_Exact][Distance]15 km[/Distance] [Direction]north[/Direction] of [Place]Kandahar[/Place] [/SpatialReference_Exact]**

Original South of Kabul

Extracted **[SpatialReference_Exact][Direction]South[/Direction] of [Place]Kabul[/Place] [/SpatialReference_Exact]**

Original near Mosul

Extracted **[SpatialReference_Vague][Direction]near[/Direction] [Place]Mosul[/Place] [/SpatialReference_Vague]**

Original in the vicinity of Kandahar

Extracted **[SpatialReference_Vague][Direction]in[/Direction] the [Place]vicinity of Kandahar[/Place] [/SpatialReference_Vague]**

6.2.10 Travel Event Extraction

The Travel rules, which are a subset of the Action rules, extract information about individuals and their travel events.

Types of Events Extracted as Travel Events

The specific types of information extracted by these rules are the following:

Type	Definition
Agent	Name of person traveling
Date	Date of travel
FromPlace	Travel origin
ToPlace	Travel destination
Travel	Movement of person

Examples

- Original In April, Pakistan President Gen. Pervez Musharraf visited Kabul.
- Extracted **[Travel_visited]**In **[Date]**April**[/Date]**, **[Agent]**Pakistan President Gen. Pervez Musharraf**[/Agent]** **[Travel]**visited**[/Travel]** **[ToPlace]**Kabul**[/ToPlace]**.**[/Travel_visited]**
- Original John Doe was arrested in December 2000, when he entered the United States from Canada.
- Extracted **[Travel_enteredFrom_When]** **[Agent]**John Doe**[/Agent]** was arrested in **[Date]**December 2000**[/Date]**, when he **[Travel]**entered**[/Travel]** **[ToPlace]**the United States**[/ToPlace]** from **[FromPlace]**Canada**[/FromPlace]**.**[/Travel_enteredFrom_When]**
- Original On Christmas Day, Queen Elizabeth arrived in San Francisco from England.
- Extracted **[Travel_arrivedInFrom]**On **[Date]**Christmas Day**[/Date]**, **[Agent]**Queen Elizabeth**[/Agent]** **[Travel]**arrived**[/Travel]** in **[ToPlace]**San Francisco**[/ToPlace]** from **[FromPlace]**England**[/FromPlace]**.**[/Travel_arrivedInFrom]**
- Original On Dec. 1, Clinton flew back to the United States from Germany.
- Extracted **[Travel_cameToFrom]**On **[Date]**Dec. 1**[/Date]**, **[Agent]**Clinton**[/Agent]** **[Travel]**flew**[/Travel]** back to **[ToPlace]**the United States**[/ToPlace]** from **[FromPlace]**Germany**[/FromPlace]**.**[/Travel_cameToFrom]**
- Original On Christmas, Mohamed departed Germany for the Netherlands.
- Extracted **[Travel_departedFor]**On **[Date]**Christmas**[/Date]**, **[Agent]**Mohamed**[/Agent]** **[Travel]**departed**[/Travel]** **[FromPlace]**Germany**[/FromPlace]** for **[ToPlace]**the Netherlands**[/ToPlace]**.**[/Travel_departedFor]**
- Original In 1992, Mohamad Hammoud gained entry to the U.S. through Venezuela.
- Extracted **[Travel_gainedEntryIntoFrom]**In **[Date]**1992**[/Date]**, **[Agent]**Mohamad Hammoud**[/Agent]** **[Travel]**gained entry**[/Travel]** to **[ToPlace]**the U.S.**[/ToPlace]** through **[FromPlace]**Venezuela**[/FromPlace]**.**[/Travel_gainedEntryIntoFrom]**
- Original On Oct. 10, Hambali crossed into Thailand from Laos.
- Extracted **[Travel_crossedIntoFrom]**On **[Date]**Oct. 10**[/Date]**, **[Agent]**Hambali**[/Agent]** **[Travel]**crossed**[/Travel]** into **[ToPlace]**Thailand**[/ToPlace]** from **[FromPlace]**Laos**[/FromPlace]**.**[/Travel_crossedIntoFrom]**
- Original In April, Mr. Singh crossed over into Pakistan from Afghanistan.

Extracted **[Travel_crossedOverToFrom]**In **[Date]**April^[/Date], **[Agent]**Mr. Singh^[/Agent]
[Travel]crossed over^[/Travel] into **[ToPlace]**Pakistan^[/ToPlace] from
[FromPlace]Afghanistan^[/FromPlace].^[/Travel_crossedOverToFrom]

Original Mr. Singh was arrested in April, when he crossed over into Pakistan from Afghanistan.

Extracted **[Travel_crossedOverToFrom_When]****[Agent]**Mr. Singh^[/Agent] was arrested in
[Date]April^[/Date], when he **[Travel]**crossed over^[/Travel] into
[ToPlace]Pakistan^[/ToPlace] from **[FromPlace]**Afghanistan^[/FromPlace].^[/Travel_crossedOverToFrom_When]

Original In 1980, Adnan Al-Ghoul escaped to Lebanon from Israel.

Extracted **[Travel_escapedToFrom]**In **[Date]**1980^[/Date], **[Agent]**Adnan Al-Ghoul^[/Agent]
[Travel]escaped^[/Travel] to **[ToPlace]**Lebanon^[/ToPlace] from
[FromPlace]Israel^[/FromPlace].^[/Travel_escapedToFrom]

Original Broderick was arrested on March 14, 2002 when he fled for Canada at the International Bridge.

Extracted **[Travel_fledFor_When]****[Agent]**Broderick^[/Agent] was arrested on **[Date]**March
14, 2002^[/Date] when he **[Travel]**fled^[/Travel] for **[ToPlace]**Canada^[/ToPlace]
[ToPlace] at the International Bridge.^[/Travel_fledFor_When]

Original In 1998, Abdulla Ocalan attempted to enter Italy.

Extracted **[Travel_attemptedToEnter]**In **[Date]**1998^[/Date], **[Agent]**Abdulla Ocalan^[/Agent]
[Agent] **[Travel]**attempted to enter^[/Travel] **[ToPlace]**Italy^[/ToPlace].^[/Travel_attemptedToEnter]

6.3 Enterprise Fact Extraction

The specialized enterprise content includes sets of rules that address domain-specific extraction requirements for an enterprise.

You can use this enterprise content to extract these specific types of information when processing and analyzing text:

Rule Set Description

- Management Changes
Extracts information related to a change in an individual's title and company, including any information about the previous or future title holder.
-

Rule Set Description

- **Membership Information**
Extracts personal membership information about an individual, and the position held within the organization.

- **Mergers and Acquisitions**
Extracts information about mergers and acquisitions

- **Organizational Information**
Extracts information about an organization, such as founder, location, or contact information

- **Product Releases**
Extracts information about product releases

i Note

The enterprise content is included in and supports the English language module only.

For details about using enterprise content to enhance extraction rules, refer to the *SAP Data Services Text Data Processing Extraction Customization Guide*.

For a full list of entity and subentity names, see [Entity Type Names and Subentity Type Names For Fact Extraction \[page 378\]](#) and note those whose name begins with "Enterprise".

Related Information

[Membership Information Extraction \[page 371\]](#)

[Management Change Event Extraction \[page 368\]](#)

[Product Release Event Extraction \[page 376\]](#)

[Merger Information Extraction \[page 373\]](#)

[Organizational Information Extraction \[page 356\]](#)

6.3.1 Management Change Event Extraction

Extracts information related to a change in an individual's title and company, including any information about the previous or future title holder.

The management changes rules file includes these two groups of rules:

Types of Events Extracted as Management Change Events

HireEvent	These rules extract patterns related to the start of employment or promotion
ResignEvent	These rules extract patterns related to the end of employment

Information Types

The specific types of information extracted by these rules are the following:

Type	Definition
Organization	Organization doing the hiring, firing or promoting
Action	Type of event (appointment, resignation, demotion, termination, and so on)
Person	Employee being hired/promoted or resigning/being terminated
Position	Title associated with employee

Original Brad Jones joins CooperVision's management team as vice president of U.S. sales.

Extracted **[HireEvent]** **[Person]**Brad Jones**[/Person]** **[Action]**joins**[/Action]**
[Organization]CooperVision**[/Organization]**'s management team as
[Position]vice president of U.S. sales**[/Position]** **[/HireEvent]**.

Original Archer Daniels Midland named Juan Luciano as chief executive, effective Jan. 1.

Extracted **[HireEvent]** **[Organization]**Archer Daniels Midland**[/Organization]**
[Action]named**[/Action]** **[Person]**Juan Luciano**[/Person]** as **[Position]**chief
executive**[/Position]** **[/HireEvent]**, effective Jan. 1.

Original Banco Santander has named former CFO Jose Antonio Alvarez as chief executive officer to replace Javier Marin.

Extracted **[HireEvent]** **[Organization]**Banco Santander**[/Organization]** has
[Action]named**[/Action]** former CFO **[Person]**Jose Antonio Alvarez**[/Person]**
as **[Position]**chief executive officer**[/Position]** **[/HireEvent]** to replace
Javier Marin.

Original Big Bear Networks, delivering intelligence into the Optical/Electrical interface for enterprise, metro and carrier networks, today named Amit Jain as its president and chief executive officer.

Extracted **[HireEvent][Organization]**Big Bear Networks**[/Organization]**, delivering intelligence into the Optical/Electrical interface for enterprise, metro and carrier networks, today **[Action]**named**[/Action]** **[Person]**Amit Jain**[/Person]** as its **[Position]**president and chief executive officer**[/Position]** **[/HireEvent]**.

Original United Technologies Corp's appointment of Chief Financial Officer Gregory Hayes as chief executive is providing investors some comfort.

Extracted **[HireEvent][Organization]**United Technologies Corp**[/Organization]**'s **[Action]**appointment**[/Action]** of Chief Financial Officer **[Person]**Gregory Hayes**[/Person]** as **[Position]**chief executive**[/Position]** **[/HireEvent]** is providing investors some comfort.

Original John Scheck ended up being hired by Mr. E.J. Ridings at Trump Mortgage as the company's national sales manager.

Extracted **[HireEvent][Person]**John Scheck**[/Person]** ended up being **[Action]**hired**[/Action]** by Mr. E.J. Ridings at **[Organization]**Trump Mortgage**[/Organization]** as the company's **[Position]**national sales manager**[/Position]** **[/HireEvent]**.

Original PepsiCo President Zein Abdalla, widely viewed as a potential successor to Chairman and CEO Indra Nooyi, is leaving the company Dec. 31.

Extracted **[ResignEvent][Organization]**PepsiCo**[/Organization]** **[Position]**President**[/Position]** **[Person]**Zein Abdalla**[/Person]**, widely viewed as a potential successor to Chairman and CEO Indra Nooyi, is **[Action]**leaving**[/Action]** **[/ResignEvent]** the company Dec. 31.

Original Sanofi says Christopher Viehbacher] has resigned as CEO after its board unanimously decided to remove him.

Extracted **[ResignEvent][Organization]**Sanofi**[/Organization]** says **[Person]**Christopher Viehbacher**[/Person]** has **[Action]**resigned**[/Action]** as **[Position]**CEO**[/Position]** after its board unanimously decided to remove him.**[/ResignEvent]**

Original Southern Community Financial Corporation (Nasdaq: SCMF; SCMFO) announced today that on February 5, 2005, Richard M. Cobb, Executive Vice President, Chief Operating Officer and Chief Financial Officer of Southern Community Financial Corporation, announced his resignation from the Company effective today.

Extracted **[ResignEvent][Organization]**Southern Community Financial Corporation**[/Organization]** (Nasdaq: SCMF; SCMFO) announced today that on February 5, 2005, **[Person]**Richard M. Cobb**[/Person]**, **[Position]**Executive Vice President, Chief Operating Officer and Chief Financial Officer of Southern Community Financial Corporation**[/Position]**, announced his **[Action]**resignation**[/Action]****[/ResignEvent]** from the Company effective today.

Original Mr. Victor Oppleman resigned as president of MainNerve, Inc. effective March 4, 2005.

Extracted **[ResignEvent][Person]**Mr. Victor Oppleman**[/Person]** **[Action]**resigned**[/Action]** as **[Position]**president**[/Position]** of **[Organization]**MainNerve, Inc.**[/Organization]****[/ResignEvent]** effective March 4, 2005.

Original Last month, Gap said Mr. Murphy would step down as CEO at the company, where he reversed a long-running sales slump but more recently has struggled to reinvigorate its eponymous brand.

Extracted Last month, **[ResignEvent][Organization]**Gap**[/Organization]** said **[Person]**Mr. Murphy**[/Person]** would **[Action]**step down**[/Action]** as **[Position]**CEO**[/Position]** at the company, where he reversed a long-running sales slump but more recently has struggled to reinvigorate its eponymous brand.**[/ResignEvent]**

6.3.2 Membership Information Extraction

The member rules are designed to extract personal membership information about an individual, and the position held within the organization.

Information types

The specific types of information extracted by these rules are the following:

Type	Definition
Organization	Organization to which a person belongs
Person	Employee or associate of an organization
Position	Person's title or association within an organization

Examples

- Original The new Chief Operating Officer of Santander (NYSE:SAN) is Jose Garcia Cantera.
- Extracted The new **[Member][Position]**Chief Operating Officer**[/Position]** of **[Organization]**Santander**[/Organization]** (NYSE:SAN) is **[Person]**Jose Garcia Cantera**[/Person]**. **[/Member]**
- Original Southern Community Financial Corporation announced the appointment of Richard M. Cobb, Executive Vice President, Chief Operating Officer and Chief Financial Officer!
- Extracted **[Member][Organization]**Southern Community Financial Corporation**[/Organization]** announced the appointment of **[Person]**Richard M. Cobb**[/Person]**, **[Position]**Executive Vice President, Chief Operating Officer and Chief Financial Officer**[/Position]** **[/Member]**!
- Original Archer Daniels Midland named Juan Luciano as chief executive, effective Jan. 1.
- Extracted **[Member][Organization]**Archer Daniels Midland**[/Organization]** named **[Person]**Juan Luciano**[/Person]** as **[Position]**chief executive**[/Position]** **[/Member]**, effective Jan. 1.
- Original Darden Chief Financial Officer Brad Richmond to retire amidst leadership changes.
- Extracted **[Member][Organization]**Darden**[/Organization]** **[Position]**Chief Financial Officer**[/Position]** **[Person]**Brad Richmond**[/Person]** **[/Member]** to retire amidst leadership changes.
- Original Andi Owen, 49, who leads the Gap outlet division, will resign as global president from Banana Republic.
- Extracted **[Member][Person]**Andi Owen**[/Person]**, 49, who leads the Gap outlet division, will resign as **[Position]**global president**[/Position]** from **[Organization]**Banana Republic**[/Organization]** **[/Member]**.
- Original Mark Carney named next head of the Bank of England.
- Extracted **[Member][Person]**Mark Carney**[/Person]** named **[Position]**next head**[/Position]** of the **[Organization]**Bank of England**[/Organization]** **[/Member]**.
- Original Hemisphere Engineering, founded 57 years ago by its CEO John Chomiak, is in advanced negotiations with Vancouver-based MCW Group of Companies.

Extracted **[Member] [Organization]Hemisphere Engineering[/Organization],** founded 57 years ago by its **[Position]CEO[/Position] [Person]John Chomiak[/Person][Member],** is in advanced negotiations with Vancouver-based MCW Group of Companies.

Original "Philip Morris is accounting for the cloud over the market," said Alan Ackerman, a market strategist at Fahnestock & Co!

Extracted "Philip Morris is accounting for the cloud over the market," said **[Member] [Person]Alan Ackerman[/Person],** a **[Position]market strategist[/Position] at [Organization]Fahnestock & Co[/Organization] [Member]!**

6.3.3 Merger Information Extraction

Extracts information about mergers and acquisitions.

Information Types

The specific types of information extracted by these rules are the following:

Type	Definition
Action	Type of transaction (merger, takeover, acquisition)
OrganizationA	Organization doing the buying or acquiring
OrganizationB	Organization being bought or acquired
Price	Price associated with the transaction (optional)

Original American Medicare Corp said Thursday it signed a letter of intent to merge into InfoCure Corp.

Extracted **[MergeEvent] [OrganizationA]American Medicare Corp[/OrganizationA]** said Thursday it signed a letter of intent to **[Action]merge[/Action]** into **[OrganizationB]InfoCure Corp. [/OrganizationB] [MergeEvent]**

Original Blue Cross and Blue Shield has agreed to merge with Anthem Inc, people familiar with the situation said.

Extracted **[MergeEvent] [OrganizationA]Blue Cross and Blue Shield[/OrganizationA]** has agreed to **[Action]merge[/Action]** with **[OrganizationB]Anthem Inc[/OrganizationB] [MergeEvent],** people familiar with the situation said.

Original Rolls-Royce plc to acquire full control of Rolls-Royce Power Systems for \$1.2 billion.

Extracted **[BuyEvent] [OrganizationA]**Rolls-Royce plc**[/OrganizationA]** to **[Action]**acquire**[/Action]** full control of **[OrganizationB]**Rolls-Royce Power Systems**[/OrganizationB]** for **[Price]**\$1.2 billion**[/Price]** **[/BuyEvent]**.

Original United Dominion Realty Trust Inc said it agreed to acquire South West Property Trust for \$312 million in stock.

Extracted **[BuyEvent] [OrganizationA]**United Dominion Realty Trust Inc**[/OrganizationA]** said it agreed to **[Action]**acquire**[/Action]** **[OrganizationB]**South West Property Trust**[/OrganizationB]** for **[Price]**\$312 million in stock**[/Price]** **[/BuyEvent]**.

Original YouTube division announced yesterday that it had acquired Directr.

Extracted **[BuyEvent] [OrganizationA]**YouTube**[/OrganizationA]** division announced yesterday that it had **[Action]**acquired**[/Action]** **[OrganizationB]**Directr**[/OrganizationB]** **[/BuyEvent]**.

Original Symix Systems Inc Friday said it acquired Canadian-based Visual Applications Software Inc for an undisclosed sum.

Extracted **[BuyEvent] [OrganizationA]**Symix Systems Inc**[/OrganizationA]** Friday said it **[Action]**acquired**[/Action]** **[OrganizationB]**Canadian-based Visual Applications Software Inc**[/OrganizationB]** for an **[Price]**undisclosed sum**[/Price]** **[/BuyEvent]**.

Original RCA stock shot up just prior to an announcement that the company would be merging with General Electric.

Extracted **[MergeEvent] [OrganizationA]**RCA**[/OrganizationA]** stock shot up just prior to an announcement that the company would be **[Action]**merging**[/Action]** with **[OrganizationB]**General Electric**[/OrganizationB]** **[/MergeEvent]**.

6.3.4 Organizational Information Extraction

Extracts core information about companies and organizations.

Information Types

The specific types of information extracted by these rules include the organization name and at least one of the following:

Type	Definition
Address	Street address of the organization
Email	E-mail address
Fax	Fax number, with extension
Location	Location of the organization
Organization	Organization about which the information is collected
Person	Individual to contact, including title and department where provided
Phone	Phone number, with extension
Url	Web Address

Examples

Original Patti J. McAtee, Director of CalEnergy, 402-341-4500.

Extracted **[OrganizationInfo][Person]**Patti J. McAtee**[/Person]**, Director of **[Organization]**CalEnergy**[/Organization]**, **[Phone]**402-341-4500**[/Phone]****[/OrganizationInfo]**.

Original Marriott International press releases are available through Company News On-Call by fax, 800-758-5804, ext. 532963.

Extracted **[OrganizationInfo][Organization]**Marriott International**[/Organization]** press releases are available through Company News On-Call by fax, **[Fax]**800-758-5804, ext. 532963**[/Fax]****[/OrganizationInfo]**.

Original Inxight Software Inc., 500 Macara Avenue, Sunnyvale, CA 94085, U.S.A., Email: info@inxight.com

Extracted **[OrganizationInfo][Organization]**Inxight Software Inc.**[/Organization]**, **[Address]**500 Macara Avenue, Sunnyvale, CA 94085, U.S.A.**[/Address]**, Email: **[Email]**info@inxight.com**[/Email]****[/OrganizationInfo]**

Original New York-based American Express Co. is best known for its credit card.

Extracted **[OrganizationInfo][Location]**New York-based**[/Location]** **[Organization]**American Express Co.**[/Organization]****[/OrganizationInfo]** is best known for its credit card.

Original Michelin Tyre is a unit of France's Michelin S.A.

Extracted Michelin Tyre is a unit of **[OrganizationInfo]** **[Nationality]**France**[/Nationality]**'s **[Organization]**Michelin S.A.**[/Organization]****[/OrganizationInfo]**

Original Microsoft (Nasdaq:MSFT) announced this quarter's earnings.

Extracted **[OrganizationInfo]** **[Organization]**Microsoft**[/Organization]** (**[Ticker]**Nasdaq:MSFT**[/Ticker]**) **[/OrganizationInfo]** announced this quarter's earnings.

Original Kodak shares were up 4.5 percent to \$45.87.

Extracted **[OrganizationInfo]** **[Organization]**Kodak**[/Organization]** shares were up 4.5 percent to **[Price]**\$45.87**[/Price]** **[/OrganizationInfo]**.

6.3.5 Product Release Event Extraction

Extracts information about the announcement of new products, including the company, date, and price.

The rules try to isolate novel product names and do not rely on the established set of extraction-recognized PRODUCT entities.

Information Types

The specific types of information extracted by these rules are the following:

Type	Definition
Action	Type of event
Date	Date associated with product release or announcement thereof (optional)
Organization	Organization releasing a product
Price	Price associated with the product (optional)
Product	Product being released

The following table describes the rules available for product release:

Original iPerceptions, Inc. (www.iperceptions360.com), an Application Service Provider (ASP) based in New York with offices in Montreal, launched a unique optimizing solution, WebValidator.

Extracted **[ProductRelease][Organization]**iPerceptions, Inc.**[/Organization]** (www.iperceptions360.com), an Application Service Provider (ASP) based in New York with offices in Montreal, **[Date]**recently**[/Date]** **[Action]**launched**[/Action]** a unique optimizing solution, **[Product]**WebValidator**[/Product]** **[/ProductRelease]**.

Original TRW Automotive Holdings Corp. announced plans to unveil the company's integrated safety system platform.

Extracted **[ProductRelease][Organization]**TRW Automotive Holdings Corp.**[/Organization]** announced plans to **[Action]**unveil**[/Action]** the company's **[Product]**integrated safety system platform**[/Product]** **[/ProductRelease]**.

Original Apple (R) today introduced the iPod(R) nano, a revolutionary full-featured iPod that holds 1,000 songs yet is thinner than a standard #2 pencil.

Extracted **[ProductRelease][Organization]**Apple**[/Organization]** (R) **[Date]**today**[/Date]** **[Action]**introduced**[/Action]** the **[Product]**iPod(R) nano**[/Product]** **[/ProductRelease]**, a revolutionary full-featured iPod that holds 1,000 songs yet is thinner than a standard #2 pencil.

Original Citect announces Switch2Citect, an automated conversion tool for SCADA users to upgrade to CitectSCADA from one or multiple legacy systems with minimal cost and effort.

Extracted **[ProductRelease][Organization]**Citect**[/Organization]** **[Action]**announces**[/Action]** **[Product]**Switch2Citect**[/Product]** **[/ProductRelease]**, an automated conversion tool for SCADA users to upgrade to CitectSCADA from one or multiple legacy systems with minimal cost and effort.

Original Compaq Computer Corp. Monday unveiled a new line of Presario home computers, including a couple of units priced under \$1,000 with several features aimed at turning the PC into a consumer appliance.

Extracted **[ProductRelease][Organization]**Compaq Computer Corp.**[/Organization]** **[Date]**Monday**[/Date]** **[Action]**unveiled**[/Action]** a **[Product]**new line of Presario home computers**[/Product]**, including a couple of units priced under **[Price]**\$1,000**[/Price]** **[/ProductRelease]** with several features aimed at turning the PC into a consumer appliance.

Original Microsoft Corp. (Nasdaq: MSFT) today announced the availability of Microsoft(R) Investor 2.0, a new version of its online investing service designed to help people organize and understand their personal investments and investment decisions.

Extracted **[ProductRelease]** **[Organization]**Microsoft Corp.**[/Organization]** (Nasdaq: MSFT) **[Date]**today**[/Date]** announced the **[Action]**availability**[/Action]** of **[Product]**Microsoft(R) Investor 2.0**[/Product]** **[/ProductRelease]**, a new version of its online investing service designed to help people organize and understand their personal investments and investment decisions.

Original Seagate Technology's Seagate Software said on Thursday it launched Seagate ExecView, touting it as the first solution that provides integrated, cross-platform storage management for enterprise networks running Microsoft Windows NT and Novell NetWare operating systems

Extracted **[ProductRelease]** **[Organization]**Seagate Technology**[/Organization]**'s Seagate Software said on **[Date]**Thursday**[/Date]** it **[Action]**launched**[/Action]** **[Product]**Seagate ExecView**[/Product]** **[/ProductRelease]**, touting it as the first solution that provides integrated, cross-platform storage management for enterprise networks running Microsoft Windows NT and Novell NetWare operating systems.

6.4 Entity Type Names and Subentity Type Names For Fact Extraction

Lists all the names of entity types and subentity types for Fact Extraction.

The following table lists all the possible entity and subentity values that can be displayed in the output from Fact Extraction rules and dictionaries, listed by “family” of Fact Extraction.

In this table, you can filter on strings in one or more of the three columns. You can filter to show only rows that contain a given string in any column. You can sort on the **Family** and the **Name of Entity or Subentity** columns.

Ffrom

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 0 0 1	VoC-Sentiment	Sentiment	A span of text referring to sentiments or problems, including the strength of the sentiment, such as strong or weak, and optionally a topic (what the sentiment/problem is about)
0 0 0 2	VoC-Sentiment	StrongPositiveSentiment	A strong positive opinion, such as "great" or "excellent"

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 0 0 3	VoC-Sentiment	WeakPositiveSentiment	A weak positive opinion, such as "nice" or "fine"
0 0 0 4	VoC-Sentiment	NeutralSentiment	An opinion that expresses ambivalence, such as "OK" or "acceptable"
0 0 0 5	VoC-Sentiment	WeakNegativeSentiment	A weak negative opinion, such as "bad" or "dislike"
0 0 0 6	VoC-Sentiment	StrongNegativeSentiment	A strong negative opinion, such as "hate" or "terrible"
0 0 0 7	VoC-Sentiment	MinorProblem	An opinion describing an impediment the customer can work around, such as "too slow" or "buggy"
0 0 0 8	VoC-Sentiment	MajorProblem	An opinion describing an impediment the customer cannot work around, such as "incomprehensible" or "defective"
0 0 0 9	VoC-Sentiment	Topic	The subject of the sentiment, answering the question "What is it that the customer is expressing their feelings about?"
0 10 1	VoC-Emoticon	Emoticon	A span of text expressing a person's feeling or mood, containing a pictorial representation (including emoji)
0 10 2	VoC-Emoticon	StrongPositiveEmoticon	Emoticon conveying strong positive sentiment
0 10 3	VoC-Emoticon	WeakPositiveEmoticon	Emoticon conveying weak positive sentiment
0 10 4	VoC-Emoticon	NeutralEmoticon	Emoticon conveying neutral sentiment
0 10 5	VoC-Emoticon	WeakNegativeEmoticon	Emoticon conveying weak negative sentiment

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 10 6	VoC-Emoticon	StrongNegativeEmoticon	Emoticon conveying strong negative sentiment
0 2 0 1	VoC-Request	Request	Span of text expressing a customer's wish for a change or enhancement or to be contacted
0 2 0 2	VoC-Request	GeneralRequest	A request for an enhancement or for something new, such as "please add", "please create", or "would like"
0 2 0 3	VoC-Request	ContactRequest	A request for direct or immediate contact, such as "please contact me" or "call me"
0 2 0 4	VoC-Request	Topic	The subject of the request, answering the question "What is the request about?"
0 3 0 1	VoC-Profanity	AMBIGUOUSPROFANITY	Words and phrases that are pejorative only in certain contexts
0 3 0 2	VoC-Profanity	UNAMBIGUOUSPROFANITY	Words and phrases that are always pejorative, regardless of the context
0 4 0 1	PublicSector-Common	COMMON_ADDRESS1 & COMMON_ADDRESS2	Common nouns for addresses, mirroring their named counterparts even if there is not any real difference in content between them
0 4 0 2	PublicSector-Common	COMMON_CONTINENT	Common nouns for the entirety of any continent
0 4 0 3	PublicSector-Common	COMMON_COUNTRY	Common nouns for the entirety of any country, including disputed territories
0 4 0 4	PublicSector-Common	COMMON_FACILITY@AIRPORT	Common nouns for air transportation terminals

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 4 0 5	PublicSector-Common	COMMON_FACILITY@BUILDGROUPS	Common nouns for architectural and civil engineering structures, and outdoor spaces
0 4 0 6	PublicSector-Common	COMMON_FACILITY@PATH	Common nouns for man-made structures that allow fluids, energy, persons, animals, or vehicles to pass from one location to another
0 4 0 7	PublicSector-Common	COMMON_FACILITY@PLANT	Common nouns for man-made structures used for industrial purposes
0 4 0 8	PublicSector-Common	COMMON_FACILITY@SUBAREA	Common nouns for portions of man-made facilities able to contain people, animals, or objects
0 4 0 9	PublicSector-Common	COMMON_GEO_AREA@DOMESTIC	Common nouns for non-political geographical areas capturing a significant land mass that do not cross national borders
0 41 0	PublicSector-Common	COMMON_GEO_AREA@INTL	Common nouns for non-political geographical areas capturing a significant land mass that do cross international borders
0 41 1	PublicSector-Common	COMMON_GEO_FEATURE@BOUNDARY	Common nouns for non-political geographical locations such as a border
0 41 2	PublicSector-Common	COMMON_GEO_FEATURE@CELESTIAL	Common nouns for non-political geographical locations outside of Earth, such as planets
0 41 3	PublicSector-Common	COMMON_GEO_FEATURE@LAND	Common nouns for non-political geographical locations: bodies of land, such as deserts, mountains, volcanoes, plateaus, plains, peninsulas, canyons, beaches, and so on
0 41 4	PublicSector-Common	COMMON_GEO_FEATURE@WATER	Common nouns for non-political geographical locations: bodies of water
0 41 5	PublicSector-Common	COMMON_LOCALITY	Common nouns for cities
0 41 6	PublicSector-Common	COMMON_ORGANIZATION@COMMERCIAL	Common nouns for companies

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 41 7	PublicSector-Common	COMMON_ORGANIZATION@EDUCATIONAL	Common nouns for institutions focused on education
0 41 8	PublicSector-Common	COMMON_ORGANIZATION@ENTERTAINMENT	Common nouns for institutions focused on entertainment
0 41 9	PublicSector-Common	COMMON_ORGANIZATION@GOVERNMENT	Common nouns for institutions related to government, politics, or the state
0 4 2 0	PublicSector-Common	COMMON_ORGANIZATION@MEDIA	Common nouns for institutions related to the media
0 4 21	PublicSector-Common	COMMON_ORGANIZATION@MEDICALSCIENCE	Institutions related to medicine or research
0 4 2 2	PublicSector-Common	COMMON_ORGANIZATION@OTHER	Common nouns for organizations that do not fit into a more specific subtype
0 4 2 3	PublicSector-Common	COMMON_ORGANIZATION@RELIGIOUS	Common nouns for institutions related to religion
0 4 2 4	PublicSector-Common	COMMON_ORGANIZATION@SPORTS	Common nouns for institutions related to sports
0 4 2 5	PublicSector-Common	COMMON_PERSON	Common nouns for persons
0 4 2 6	PublicSector-Common	COMMON_PRECURSOR@CHEMICAL	Common names of chemical substances that are used in the manufacturing of weapons
0 4 2 7	PublicSector-Common	COMMON_PRECURSOR@NUCLEAR	Commons names of nuclear and radiological substances used in the manufacturing of weapons
0 4 2 8	PublicSector-Common	COMMON_REGION@MAJOR	Common nouns of administrative units above the city and below the country: states and provinces

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 4 2 9	PublicSector-Common	COMMON_REGION@MINOR	Common nouns of administrative units above the city and below the country: counties and districts
0 4 3 0	PublicSector-Common	COMMON_VEHICLE@AIR	Common names for air vehicles
0 4 31	PublicSector-Common	COMMON_VEHICLE@LAND	Common names for land vehicles
0 4 3 2	PublicSector-Common	COMMON_VEHICLE@OTHER	Common names for vehicles that do not fit into a more specific subtype
0 4 3 3	PublicSector-Common	COMMON_VEHICLE@SUBAREA	Common names for portions of vehicles in which humans can fit
0 4 3 4	PublicSector-Common	COMMON_VEHICLE@WATER	Common names for water vehicles
0 4 3 5	PublicSector-Common	COMMON_WEAPON@BIOLOGICAL	Common names for biological weapons
0 4 3 6	PublicSector-Common	COMMON_WEAPON@BLUNT	Common names for conventional weapons designed to bludgeon
0 4 3 7	PublicSector-Common	COMMON_WEAPON@CHEMICAL	Common names for chemical weapons
0 4 3 8	PublicSector-Common	COMMON_WEAPON@EXPLODING	Common names for conventional weapons designed to explode
0 4 3 9	PublicSector-Common	COMMON_WEAPON@NUCLEAR	Common names for nuclear weapons

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 4 4 0	PublicSector-Common	COMMON_WEAPON@OTHER	Common names for conventional weapons that do not fit into a more specific subtype
0 4 41	PublicSector-Common	COMMON_WEAPON@PROJECTILE	Common names for conventional weapons designed to be projected at great speed
0 4 4 2	PublicSector-Common	COMMON_WEAPON@SHARP	Common names for conventional weapons designed to cut
0 4 4 3	PublicSector-Common	COMMON_WEAPON@SHOOTING	Common names for conventional weapons designed to send projectile objects at great speed
0 5 0 1	PublicSector-Action	Action_Buy_Active	A sentence or clause expressing a "Buy" event involving persons or organizations (active voice)
0 5 0 2	PublicSector-Action	Action_Buy_Passive	A sentence or clause expressing a "Buy" event involving persons or organizations (passive voice)
0 5 0 3	PublicSector-Action	Action_Capture_Active	A sentence or clause expressing a "Capture" event involving persons or organizations (active voice)
0 5 0 4	PublicSector-Action	Action_Capture_Passive	A sentence or clause expressing a "Capture" event involving persons or organizations (passive voice)
0 5 0 5	PublicSector-Action	Action_Command_Active	A sentence or clause expressing a "Command" event involving persons or organizations (active voice)
0 5 0 6	PublicSector-Action	Action_Command_Appositive	A sentence or clause expressing a "Command" event involving persons or organizations (apposition)
0 5 0 7	PublicSector-Action	Action_Command_Passive	A sentence or clause expressing a "Command" event involving persons or organizations (passive voice)

or
di
n
al

Family	Name of Entity (←) or Subentity (→)	Description
0508 PublicSector-Action	Action_Communicate_Meet_Active	A sentence or clause expressing a "Meet" event involving persons or organizations (active voice)
0509 PublicSector-Action	Action_Communicate_Meet_Passive	A sentence or clause expressing a "Meet" event involving persons or organizations (passive voice)
0510 PublicSector-Action	Action_Communicate_Other_Active	A sentence or clause expressing a "Communicate" event involving persons or organizations (active voice)
0511 PublicSector-Action	Action_Communicate_Other_Passive	A sentence or clause expressing a "Communicate" event involving persons or organizations (passive voice)
0512 PublicSector-Action	Action_Communicate_PhoneWrite_Active	A sentence or clause expressing a "Phone" or "Write" event involving persons or organizations (active voice)
0513 PublicSector-Action	Action_Destroy_Active	A sentence or clause expressing a "Destroy" event involving persons or organizations (active voice)
0514 PublicSector-Action	Action_Destroy_Passive	A sentence or clause expressing a "Destroy" event involving persons or organizations (passive voice)
0515 PublicSector-Action	Action_Drive_Active	A sentence or clause expressing a "Drive" event involving persons or organizations (active voice)
0516 PublicSector-Action	Action_Drive_Passive	A sentence or clause expressing a "Drive" event involving persons or organizations (passive voice)
0517 PublicSector-Action	Action_Execute_Active	A sentence or clause expressing an "Execute" event involving persons or organizations (active voice)
0518 PublicSector-Action	Action_Execute_Passive	A sentence or clause expressing an "Execute" event involving persons or organizations (passive voice)
0519 PublicSector-Action	Action_Finance_Active	A sentence or clause expressing a "Finance" event involving persons or organizations (active voice)
0520 PublicSector-Action	Action_Finance_Passive	A sentence or clause expressing a "Finance" event involving persons or organizations (passive voice)

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 5 21	PublicSector-Action	Action_Hire_Active	A sentence or clause expressing a "Hire" event involving persons or organizations (active voice)
0 5 2 2	PublicSector-Action	Action_Hire_Passive	A sentence or clause expressing a "Hire" event involving persons or organizations (passive voice)
0 5 2 3	PublicSector-Action	Action_Indict_Active	A sentence or clause expressing an "Indict" event involving persons or organizations (active voice)
0 5 2 4	PublicSector-Action	Action_Indict_Passive	A sentence or clause expressing an "Indict" event involving persons or organizations (passive voice)
0 5 2 5	PublicSector-Action	Action_Injure_Active	A sentence or clause expressing an "Injure" event involving persons or organizations (active voice)
0 5 2 6	PublicSector-Action	Action_Injure_Passive	A sentence or clause expressing an "Injure" event involving persons or organizations (passive voice)
0 5 2 7	PublicSector-Action	Action_Kill_Active	A sentence or clause expressing a "Kill" event involving persons or organizations (active voice)
0 5 2 8	PublicSector-Action	Action_Kill_Passive	A sentence or clause expressing a "Kill" event involving persons or organizations (passive voice)
0 5 2 9	PublicSector-Action	Action_Make_Active	A sentence or clause expressing a "Make" event involving persons or organizations (active voice)
0 5 3 0	PublicSector-Action	Action_Make_MakerOf	A sentence or clause expressing a "Maker" event involving persons or organizations
0 5 31	PublicSector-Action	Action_Make_Passive	A sentence or clause expressing a "Make" event involving persons or organizations (passive voice)

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 5 3 2	PublicSector-Action	Action_Participate_Active	A sentence or clause expressing a "Participate" event involving persons or organizations (active voice)
0 5 3 3	PublicSector-Action	Action_Participate_Passive	A sentence or clause expressing a "Participate" event involving persons or organizations (passive voice)
0 5 3 4	PublicSector-Action	Action_Receive_Pay_Active	A sentence or clause expressing a "Receive payment" event involving persons or organizations (active voice)
0 5 3 5	PublicSector-Action	Action_Receive_Pay_Passive	A sentence or clause expressing a "Receive payment" event involving persons or organizations (passive voice)
0 5 3 6	PublicSector-Action	Action_Survey_Active	A sentence or clause expressing a "Survey" event involving persons or organizations (active voice)
0 5 3 7	PublicSector-Action	Action_Survey_Passive	A sentence or clause expressing a "Survey" event involving persons or organizations (passive voice)
0 5 3 8	PublicSector-Action	Action_Train_Active	A sentence or clause expressing a "Train" event involving persons or organizations (active voice)
0 5 3 9	PublicSector-Action	Action_Train_Passive	A sentence or clause expressing a "Train" event involving persons or organizations (passive voice)
0 5 4 0	PublicSector-Action	Action_Train_With	A sentence or clause expressing a "Train with" event involving persons or organizations (active voice)
0 5 41	PublicSector-Action	Action_Transport_Active	A sentence or clause expressing a "Transport" event involving persons or organizations (active voice)
0 5 4 2	PublicSector-Action	Action_Transport_Passive	A sentence or clause expressing a "Transport" event involving persons or organizations (passive voice)

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 5 4 3	PublicSector-Action		Agent First of two arguments taking part in an action event
0 5 4 4	PublicSector-Action		Artifact Second of two arguments taking part in an action event (inanimate)
0 5 4 5	PublicSector-Action		Date Date expression occurring in same sentence as an action event
0 5 4 6	PublicSector-Action		Time Time expression occurring in same sentence as an action event
0 5 4 7	PublicSector-Action		Place Place expression occurring in same sentence as an action event
0 5 51	PublicSector-Action		Organization Organization taking part in an action event
0 5 5 2	PublicSector-Action		Patient Second of two arguments taking part in an action event (animate)
0 5 5 3	PublicSector-Action		Payment Payment given to Recipient in an action event
0 5 5 4	PublicSector-Action		Recipient Person or organization receiving payment in an action event
0 6 0 1	PublicSector-Military	MilitaryUnit	An expressions that refers to a military unit, often coordinated (Company C, 3rd Battalion, 67th Armor Regiment, 4th Brigade Combat Team, 101st Airborne Division)
0 6 0 2	PublicSector-Military	MILITARY_UNIT/US_AIRFORCE	Military units in the U.S. Air Force

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 6 0 3	PublicSector-Military	MILITARY_UNIT/US_ARMY	Military units in the U.S. Army
0 6 0 4	PublicSector-Military	MILITARY_UNIT/US_NAVY	Military units in the U.S. Navy
0 7 0 1	PublicSector-Organization Info	OrganizationInfo	A span of text expressing information about an organization
0 7 0 2	PublicSector-Organization Info	Address	Street address of the organization
0 7 0 3	PublicSector-Organization Info	Email	E-mail address
0 7 0 4	PublicSector-Organization Info	Fax	Fax number, with extension
0 7 0 5	PublicSector-Organization Info	Location	Location of the organization
0 7 0 6	PublicSector-Organization Info	Nationality	Nationality of the organization
0 7 0 7	PublicSector-Organization Info	Organization	Name of the organization about which the information is collected
0 7 0 8	PublicSector-Organization Info	Person	Name of individual to contact, including title and department where provided
0 7 0 9	PublicSector-Organization Info	Phone	Phone number, with extension

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 71 0	PublicSector-Organization Info		Price Price or value of the organization
0 71 1	PublicSector-Organization Info		Ticker Ticker symbol of the organization
0 71 2	PublicSector-Organization Info		Url Web Address
0 8 0 1	PublicSector-Person:Alias	PersonAlias_alias_Person	A sentence or clause containing a pattern that follows the general form "Person, alias, Person or Proper Noun, Punctuation"
0 8 0 2	PublicSector-Person:Alias	PersonAlias_alias_PersonP aren	A sentence or clause containing a pattern that follows the general form "Open Parenthesis, alias, Person or Proper Noun, Close Parenthesis"
0 8 0 3	PublicSector-Person:Alias	PersonAlias_alias_Proper	A sentence or clause containing a pattern that follows the general form "Proper Noun, alias, Person, Punctuation"
0 8 0 4	PublicSector-Person:Alias	PersonAlias_alias_ProperP aren	A sentence or clause containing a pattern that follows the general form "Proper Noun, Open Parenthesis, alias, Person, Close Parenthesis"
0 8 0 5	PublicSector-Person:Alias	PersonAlias_or	A sentence or clause containing a pattern that follows the general form "Person, Comma, or, Person, Comma"
0 8 0 6	PublicSector-Person:Alias	PersonAlias_AlsoKnownAs	A sentence or clause containing a pattern that follows the general form "Person, also known as, Person or Proper Noun"
0 8 0 7	PublicSector-Person:Alias	PersonAlias_AlsoKnownAs_W ho	A sentence or clause containing a pattern that follows the general form "Person, who is, also know as, Person or Proper Noun"
0 8 0 8	PublicSector-Person:Alias	PersonAlias_AlsoKnownAs_N Pwho	A sentence or clause containing a pattern that follows the general form "Person, Comma, NP, who is, also know as, Person or Proper Noun"
0 8 0 9	PublicSector-Person:Alias	PersonAlias_AlsoKnownAs_Q uote	A sentence or clause containing a pattern that follows the general form "Person, also known as, Open Quote, General Alias, Close Quote"

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
0 81 0	PublicSector-Person:Alias	PersonAlias_UsingTheName	A sentence or clause containing a pattern that follows the general form "Person, using the name, Person or Proper Noun"
0 81 1	PublicSector-Person:Alias	PersonAlias_UsingTheName_Who	A sentence or clause containing a pattern that follows the general form "Person, who is, using the name, Person or Proper Noun"
0 81 2	PublicSector-Person:Alias	PersonAlias_UsingTheName_NPWho	A sentence or clause containing a pattern that follows the general form "Person, Comma, NP, who is, using the name, Person or Proper Noun"
0 81 3	PublicSector-Person:Alias	PersonAlias_UsingTheName_Quote	A sentence or clause containing a pattern that follows the general form "Person, using the name, Open Quote, General Alias, CloseQuote"
0 81 4	PublicSector-Person:Alias	PersonAlias_UsingTheNames	A sentence or clause containing a pattern that follows the general form "Person, using the names, Person or Proper Noun"
0 81 5	PublicSector-Person:Alias	PersonAlias_UsingTheNames_Who	A sentence or clause containing a pattern that follows the general form "Person, who is, using the names, Person or Proper Noun"
0 81 6	PublicSector-Person:Alias	PersonAlias_UsingTheNames_NPWho	A sentence or clause containing a pattern that follows the general form "Person, Comma, NP who is, using the names, Person or Proper Noun"
0 81 7	PublicSector-Person:Alias	PersonAlias_UsingTheNames_Quote	A sentence or clause containing a pattern that follows the general form "Person, using the names, Open Quote, General Alias, CloseQuote"
0 81 8	PublicSector-Person:Alias	PersonAlias_AlsoKnownAsPRE	A sentence or clause containing a pattern that follows the general form "also known as, Person or Proper Noun, Comma, Person"
0 81 9	PublicSector-Person:Alias	PersonAlias_AlsoKnownAsPRE_Quote	A sentence or clause containing a pattern that follows the general form "also known as, Open Quote, General Alias, Close Quote, Comma, Person"
0 8 2 0	PublicSector-Person:Alias	Alias	An alias used by a person

or di n al	Family	Name of Entity (←) or Subentity (→)		Description
0 8 21	PublicSector-Person:Alias		Person	The name of a person
0 9 0 1	PublicSector-Person: Appearance	PerApp_Age		A sentence or clause linking a person and his/her age
0 9 0 2	PublicSector-Person: Appearance	PerApp_Dress		A sentence or clause linking a person and what he/she is wearing
0 9 0 3	PublicSector-Person: Appearance	PerApp_Eyes		A sentence or clause linking a person and a description of his/her eyes
0 9 0 4	PublicSector-Person: Appearance	PerApp_Hair		A sentence or clause linking a person and a description of his/her hair
0 9 0 5	PublicSector-Person: Appearance	PerApp_Height		A sentence or clause linking a person and his/her height
0 9 0 6	PublicSector-Person: Appearance	PerApp_Weight		A sentence or clause linking a person and his/her weight
0 9 0 7	PublicSector-Person: Appearance		Person	The name of a person
0 9 0 8	PublicSector-Person: Appearance		Age	Person's age
0 9 0 9	PublicSector-Person: Appearance		Dress	Person's attire
0 91 0	PublicSector-Person: Appearance		Eyes	Person's eye color
0 91 1	PublicSector-Person: Appearance		Hair	Person's hair color and possibly shape

or di n al	Family	Name of Entity (←) or Subentity (→)		Description
0 91 2	PublicSector-Person: Appearance		Height	Person's height (in numbers)
0 91 3	PublicSector-Person: Appearance		Weight	Person's weight (in numbers)
10 0 1	PublicSector-Person: Attribute	PerAtt_Address		A sentence or clause linking a person and his/her address
10 0 2	PublicSector-Person: Attribute	PerAtt_Affiliation		A sentence or clause linking a person and his/her address
10 0 3	PublicSector-Person: Attribute	PerAtt_Location		A sentence or clause linking a person and his/her affiliation
10 0 4	PublicSector-Person: Attribute	PerAtt_Location_verbSn		A sentence or clause linking a person and his/her location
10 0 5	PublicSector-Person: Attribute	PerAtt_Nationality		A sentence or clause linking a person and his/her nationality
10 0 6	PublicSector-Person: Attribute	PerAtt_Occupation		A sentence or clause linking a person and his/her occupation
10 0 7	PublicSector-Person: Attribute	PerAtt_Phone		A sentence or clause linking a person and his/her phone number
10 0 8	PublicSector-Person: Attribute	PerAtt_Phone_verbSn		A sentence or clause linking a person and his/her phone number
10 0 9	PublicSector-Person: Attribute	PerAtt_Possession		A sentence or clause linking a person and his/her possession
10 10	PublicSector-Person: Attribute	PerAtt_Vehicle		A sentence or clause linking a person and his/her vehicle
10 11	PublicSector-Person: Attribute		Person	Name of person
10 12	PublicSector-Person: Attribute		Residence	Person's address (specific or non-specific)
10 13	PublicSector-Person: Attribute		Phone	Person's phone number

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
10 14	PublicSector-Person: Attribute	Vehicle	Person's vehicle
10 15	PublicSector-Person: Attribute	Nationality	Person's nationality
10 16	PublicSector-Person: Attribute	Organization	Person's affiliation
10 17	PublicSector-Person: Attribute	Occupation	Person's occupation and/or skill
10 18	PublicSector-Person: Attribute	Artifact	Person's possession
11 0 1	PublicSector-Person: Relationship	PerRel_Associate	A sentence or clause linking a person and his/her associate(s)
11 0 2	PublicSector-Person: Relationship	PerRel_ParentChild	A sentence or clause linking a person and his/her child(ren)
11 0 3	PublicSector-Person: Relationship	PerRel_Relative	A sentence or clause linking a person and his/her relative(s)
11 0 4	PublicSector-Person: Relationship	PerRel_Sibling	A sentence or clause linking a person and his/her sibling(s)
11 0 5	PublicSector-Person: Relationship	PerRel_Spouse	A sentence or clause linking a person and his/her spouse
11 0 6	PublicSector-Person: Relationship	Associate	Name of person's associate
11 0 7	PublicSector-Person: Relationship	Child	Name of person's child
11 0 8	PublicSector-Person: Relationship	Parent	Name of person's parent
11 0 9	PublicSector-Person: Relationship	Relative	Name of person's relative
11 10	PublicSector-Person: Relationship	Sibling	Name of person's sibling
11 11	PublicSector-Person: Relationship	Spouse	Name of person's spouse

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
12 0 1	PublicSector-Spatial Reference	SpatialReference_Exact	A sentence or clause linking a location with a spatial reference (50 miles west of Boston)
12 0 2	PublicSector-Spatial Reference	SpatialReference_Vague	A sentence or clause linking a location with a vague spatial reference (near Boston; in the vicinity of Paris)
12 0 3	PublicSector-Spatial Reference	Direction	Keyword indicating a direction (north, near, not too far from, and so on)
12 0 4	PublicSector-Spatial Reference	Distance	A distance in km, miles
12 0 5	PublicSector-Spatial Reference	Place	Name of a place used in the relative spatial reference
13 0 3	PublicSector-Travel	Travel_arrivedIn_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, arrived, in, TargetDestinations, from, Source Destination"
13 0 4	PublicSector-Travel	Travel_arrivedInFrom	A sentence or clause containing a pattern that follows the general form "Person, arrived, in, TargetDestinations, from, Source Destination"
13 0 5	PublicSector-Travel	Travel_attemptedEntryIntoFrom	A sentence or clause containing a pattern that follows the general form "Person, attempted entry, to, Target Destinations, from, Source Destination"
13 0 6	PublicSector-Travel	Travel_attemptedEntryIntoFrom_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, attempted entry, to, Target Destinations, from, Source Destination"
13 0 7	PublicSector-Travel	Travel_attemptedToEnter	A sentence or clause containing a pattern that follows the general form "Person, attempted to enter, Target Destinations"
13 0 8	PublicSector-Travel	Travel_attemptedToEnter_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, attempted to enter, Target Destinations"

or
di
n
al

	Family	Name of Entity (←) or Subentity (→)	Description
1309	PublicSector-Travel	Travel_cameFromTo	A sentence or clause containing a pattern that follows the general form "Person, came from, Source Destination, back to, Target Destinations"
1310	PublicSector-Travel	Travel_cameFromTo_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, came, from, Source Destination, back to, Target Destinations"
1311	PublicSector-Travel	Travel_cameToFrom	A sentence or clause containing a pattern that follows the general form "Person, came back to, Target Destinations, from, Source Destination"
1312	PublicSector-Travel	Travel_cameToFrom_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, came, back to, Target Destinations, from, Source Destination"
1313	PublicSector-Travel	Travel_crossedIntoFrom	A sentence or clause containing a pattern that follows the general form "Person, crossed, the border, to, Target Destinations, from, Source Destination"
1314	PublicSector-Travel	Travel_crossedIntoFrom_When	A sentence or clause containing a pattern that follows the general form "Person, Person, Date, when, crossed, the border, to, Target Destinations, from, Source Destination"
1315	PublicSector-Travel	Travel_crossedOverFromTo	A sentence or clause containing a pattern that follows the general form "Person, crossed over, from, Source Destination, to, Target Destinations"
1316	PublicSector-Travel	Travel_crossedOverFromTo_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, crossed over, from, Source Destination, to, Target Destinations"
1317	PublicSector-Travel	Travel_crossedOverToFrom	A sentence or clause containing a pattern that follows the general form "Person, crossed over, to, Target Destinations, from, Source Destination"
1318	PublicSector-Travel	Travel_crossedOverToFrom_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, crossed over, to, Target Destinations, from, Source Destination"

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
13 19	PublicSector-Travel	Travel_departedFor	A sentence or clause containing a pattern that follows the general form "Person, departed, Source Destination, for, Target Destination"
13 2 0	PublicSector-Travel	Travel_departedFor_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, departed, Source Destination, for, Target Destinations"
13 21	PublicSector-Travel	Travel_enteredFrom	A sentence or clause containing a pattern that follows the general form "Person, entered, TargetDestinations, from, Source Destination"
13 2 2	PublicSector-Travel	Travel_enteredFrom_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, entered, TargetDestinations, from, Source Destination"
13 2 3	PublicSector-Travel	Travel_escapedFromTo	A sentence or clause containing a pattern that follows the general form "Person, Date, when, escaped, to, Target Destinations, from, Source Destination"
13 2 4	PublicSector-Travel	Travel_escapedFromTo_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, escaped, from, Source Destination, to, Target Destinations"
13 2 5	PublicSector-Travel	Travel_escapedToFrom	A sentence or clause containing a pattern that follows the general form "Person, escaped, to, Target Destinations, from, Source Destination"
13 2 6	PublicSector-Travel	Travel_escapedToFrom_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, escaped, to, Target Destinations, from, Source Destination"
13 2 7	PublicSector-Travel	Travel_fledFor	A sentence or clause containing a pattern that follows the general form "Person, fled, Source Destination, for, Target Destinations"
13 2 8	PublicSector-Travel	Travel_fledFor_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, fled, Source Destination, for, Target Destinations"

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
13 2 9	PublicSector-Travel	Travel_gainedEntryIntoFrom	A sentence or clause containing a pattern that follows the general form ""Person, gained entry, to, Target Destinations, from, Source Destination"
13 3 0	PublicSector-Travel	Travel_gainedEntryIntoFrom_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, gained entry, to, Target Destinations, from, Source Destination"
13 31	PublicSector-Travel	Travel_triedToCrossIntoFrom	A sentence or clause containing a pattern that follows the general form "Person, tried to cross, the border, to, Target Destinations, from, Source Destination"
13 3 2	PublicSector-Travel	Travel_triedToCrossIntoFrom_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, tried to cross, the border, to, Target Destinations, from, Source Destination"
13 3 3	PublicSector-Travel	Travel_visited	A sentence or clause containing a pattern that follows the general form "Person, visited, Target Destinations"
13 3 4	PublicSector-Travel	Travel_visited_When	A sentence or clause containing a pattern that follows the general form "Person, Date, when, visited, Target Destinations"
13 3 5	PublicSector-Travel	Agent	First of two arguments taking part in a travel event
13 3 6	PublicSector-Travel	FromPlace	Origin of trip in a travel event
13 3 7	PublicSector-Travel	Place	Place expression occurring in same sentence as a travel event
13 3 8	PublicSector-Travel	Time	Time expression occurring in same sentence as a travel event
13 3 9	PublicSector-Travel	ToPlace	Destination in a travel event
13 4 0	PublicSector-Travel	Travel	Travel movement in a travel event

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
14 0 1	Enterprise-Management Change	HireEvent	A span of text expressing a hiring or promotion
14 0 2	Enterprise-Management Change	ResignEvent	A span of text expressing a separation from a job
14 0 3	Enterprise-Management Change	Action	Type of event (appointment, resignation, demotion, termination, an so on)
14 0 4	Enterprise-Management Change	Organization	Organization doing the hiring, firing, or promoting
14 0 5	Enterprise-Management Change	Person	Employee being hire/promoted or resigning/being terminated
14 0 6	Enterprise-Management Change	Position	Employee's title or association within an organization
15 0 1	Enterprise-Membership Info	Member	A span of text expressing a membership in an organization
15 0 2	Enterprise-Membership Info	Organization	Organization to which a person belongs
15 0 3	Enterprise-Membership Info	Person	Employee or associate of an organization
15 0 4	Enterprise-Membership Info	Position	Person's title or association within an organization
16 0 1	Enterprise-Merger Info	BuyEvent	A span of text expressing the acquisition of an organization
16 0 2	Enterprise-Merger Info	SellEvent	A span of text expressing the divestiture of an organization
16 0 3	Enterprise-Merger Info	MergeEvent	A span of text expressing the merger of organizations
16 0 4	Enterprise-Merger Info	Action	Type of transaction (merger, takeover, acquisition)

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
16 0 5	Enterprise-Merger Info	OrganizationA	Organization doing the buying or acquiring
16 0 6	Enterprise-Merger Info	OrganizationB	Organization being bought or acquired
16 0 7	Enterprise-Merger Info	Price	Price associated with the transaction (optional)
17 0 1	Enterprise-Organization Info	OrganizationInfo	A span of text expressing a statement about an organization
17 0 2	Enterprise-Organization Info	Address	Street address of the organization
17 0 3	Enterprise-Organization Info	Email	E-mail address
17 0 4	Enterprise-Organization Info	Fax	Fax number, with extension
17 0 5	Enterprise-Organization Info	Location	Location of the organization
17 0 6	Enterprise-Organization Info	Nationality	Nationality of the organization
17 0 7	Enterprise-Organization Info	Organization	Organization about which the information is collected
17 0 8	Enterprise-Organization Info	Person	Individual to contact, including title and department where provided
17 0 9	Enterprise-Organization Info	Phone	Phone number, with extension
17 10	Enterprise-Organization Info	Price	Price or value of the organization
17 11	Enterprise-Organization Info	Ticker	Ticker symbol of the organization
17 12	Enterprise-Organization Info	Url	Web address

or di n al	Family	Name of Entity (←) or Subentity (→)	Description
18 0 1	Enterprise-Product Release	ProductRelease	A span of text expressing the release of a new product
18 0 2	Enterprise-Product Release	Action	Type of event
18 0 2	Enterprise-Product Release	Organization	Organization releasing a product
18 0 3	Enterprise-Product Release	Price	Price associated with the product (optional)
18 0 4	Enterprise-Product Release	Product	Product being released
18 0 5	Enterprise-Product Release	Date	Date associated with product release or announcement thereof (optional)

Related Information



- [Sentiment Extraction \[page 315\]](#)
- [Emoticon Extraction \[page 329\]](#)
- [Request Extraction \[page 340\]](#)
- [Profanity Extraction \[page 346\]](#)
- [Common Entity Extraction \[page 348\]](#)
- [Action Event Extraction \[page 353\]](#)
- [Military Unit Extraction \[page 355\]](#)
- [Organizational Information Extraction \[page 356\]](#)
- [Person: Alias Extraction \[page 357\]](#)
- [Person: Appearance Extraction \[page 359\]](#)
- [Person: Attribute Extraction \[page 361\]](#)
- [Person: Relationship Extraction \[page 362\]](#)
- [Spatial Reference Extraction \[page 364\]](#)
- [Travel Event Extraction \[page 365\]](#)
- [Management Change Event Extraction \[page 368\]](#)
- [Merger Information Extraction \[page 373\]](#)
- [Organizational Information Extraction \[page 356\]](#)
- [Product Release Event Extraction \[page 376\]](#)

Important Disclaimers and Legal Information

Hyperlinks

Some links are classified by an icon and/or a mouseover text. These links provide additional information.

About the icons:

- Links with the icon : You are entering a Web site that is not hosted by SAP. By using such links, you agree (unless expressly stated otherwise in your agreements with SAP) to this:
 - The content of the linked-to site is not SAP documentation. You may not infer any product claims against SAP based on this information.
 - SAP does not agree or disagree with the content on the linked-to site, nor does SAP warrant the availability and correctness. SAP shall not be liable for any damages caused by the use of such content unless damages have been caused by SAP's gross negligence or willful misconduct.
- Links with the icon : You are leaving the documentation for that particular SAP product or service and are entering a SAP-hosted Web site. By using such links, you agree that (unless expressly stated otherwise in your agreements with SAP) you may not infer any product claims against SAP based on this information.

Beta and Other Experimental Features

Experimental features are not part of the officially delivered scope that SAP guarantees for future releases. This means that experimental features may be changed by SAP at any time for any reason without notice. Experimental features are not for productive use. You may not demonstrate, test, examine, evaluate or otherwise use the experimental features in a live operating environment or with data that has not been sufficiently backed up.

The purpose of experimental features is to get feedback early on, allowing customers and partners to influence the future product accordingly. By providing your feedback (e.g. in the SAP Community), you accept that intellectual property rights of the contributions or derivative works shall remain the exclusive property of SAP.

Example Code

Any software coding and/or code snippets are examples. They are not for productive use. The example code is only intended to better explain and visualize the syntax and phrasing rules. SAP does not warrant the correctness and completeness of the example code. SAP shall not be liable for errors or damages caused by the use of example code unless damages have been caused by SAP's gross negligence or willful misconduct.

Gender-Related Language

We try not to use gender-specific word forms and formulations. As appropriate for context and readability, SAP may use masculine word forms to refer to all genders.

Videos Hosted on External Platforms

Some videos may point to third-party video hosting platforms. SAP cannot guarantee the future availability of videos stored on these platforms. Furthermore, any advertisements or other content hosted on these platforms (for example, suggested videos or by navigating to other videos hosted on the same site), are not within the control or responsibility of SAP.

© 2020 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company. The information contained herein may be changed without prior notice.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

Please see <https://www.sap.com/about/legal/trademark.html> for additional trademark information and notices.