iTAG: Automatically Annotating Textual Resources with Human Intentions

Mark Kröll
Knowledge Management Institute, Graz University of Technology
Inffeldgasse 21a/II, A-8010 Graz, Austria
Email: mkroell@tugraz.at

Christian Körner
Knowledge Management Institute, Graz University of Technology
Inffeldgasse 21a/II, A-8010 Graz, Austria
Email: christian.koerner@tugraz.at

Markus Strohmaier
Knowledge Management Institute, Graz University of Technology and Know-Center
Inffeldgasse 21a/II, A-8010 Graz, Austria
Email: markus.strohmaier@tugraz.at

Abstract— **Annotations represent an increasingly popular means for organizing, categorizing and finding resources on the "social" web. Yet, only a small portion of the total resources available on the web are annotated. Work on automatic tag generation algorithms aims to tackle this problem by developing algorithms that attempt to approximate and support human tagging behavior. While existing algorithms largely focus on automatically describing the general topics covered by a resource (such as "career", "education"), we suggest focusing on a different tagging dimension: i.e. automatically annotating resources with human intentions. Intent annotations aim to describe which goals are referenced in given textual resources (such as "find a job", "get a degree"), thereby offering a new, interesting perspective on textual resources on the web. We describe a prototype – iTAG – for automatically annotating textual resources with human intent, and investigate the extent to which the automatic analysis of human intentions in textual resources is feasible. For evaluation purposes, we present results from an exploratory study that focused on annotating intent in transcripts of political speeches given by US presidential candidates in 2008.**

Index Terms—**Intent Annotation, Tagging, Automatic Tag Generation, Human Intentions, Text Understanding**

I. Introduction

Folksonomies are often characterized by a tripartite graph with hyperedges. The three disjoint, finite sets of such a graph are typically defined as 1) a set of users $u \in U$ 2) a set of resources $r \in R$ and 3) a set of annotations or tags $t \in T$ that are used by users U to annotate resources R, yielding a general model of folksonomies $F \subseteq U \times T \times R$ (cf. [9, 11, 13, 14]). "*In the wild*" represents a general model of folksonomies that is known to produce a variety of different dimensions of tags T, such as topic, time, location, author, opinion [15], sentiment [20], quality and other types of tags [2]. These dimensions are considered to be useful for a range of different purposes, due to their ability to capture information about textual resources that is not necessarily contained in the resources themselves [1]. This additional information makes annotations an increasingly popular means for organizing, categorizing and finding resources on the social web.

Yet, only a minor fraction of resources on the web are annotated [9]. This has led our research community to develop automatic tag generation algorithms aiming to augment and approximate human tagging behavior. Recent attempts include *TagAssist*, an approach to automatically suggest appropriate topic tags for blog posts (such as "politics", "news") [17] or *P-Tag*, an algorithm to automatically produce personalized tags for web pages [3]. Results reported by these early attempts are encouraging and demonstrate that for selected tagging dimensions useful approximations can be produced.

While certain dimensions of tags dominate folksonomies in many applications such as search [2], a particularly interesting yet currently not very well understood dimension of annotations is *human intent*. In contrast to *topic* or *quality* annotations, *intent* annotations focus on future states of affairs that some agent wants to achieve, and describe which goals or human intentions are relevant in the context of a given textual resource. To give an example: While a particular blog post might focus on the topics "cars" and "automobiles", the underlying intention of the author might be to "Achieve mobility" or to "Reduce ecological footprint". Intent can be assumed to play a fundamental role in user interactions on the web, including interpreting and understanding resources. Intent annotations could be useful, for example, to quickly grasp

---

Manuscript received June 16, 2010; revised August 19, 2010; accepted August 25, 2010.
Corresponding author: Mark Kröll

the main aspirations implicitly addressed by resources or to enable goal-oriented navigation of resources, such as blogs, on the web (cf. for example, [18]). Figure 1 shows an example tag cloud[1] of intent annotations.

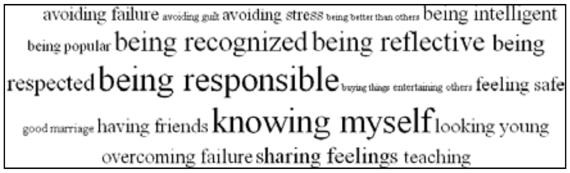

Figure 1 shows an example tag cloud of intent annotations.

Figure 1 aims to illustrate the notion of intent annotations by giving an example of a tag cloud revealing information about goals and intentions referenced in a textual resource. Without knowing the underlying resource, a range of interesting analyses becomes possible. From knowing authors' goals and interests, one might be able to infer their opinions, their relationship with other people or their attitude towards life. However, existing folksonomy-based systems do not support or encourage users in assigning intent tags and as a result - this type of annotation is hardly used "in the wild".

In this work, we study the extent to which it is feasible to *automatically annotate textual resources with human intentions*. The paper is structured as follows: First, we provide an overview of related work on tagging and algorithms for automatic tag generation. In Section III, we briefly describe and characterize the task of Intent Annotation. Section IV introduces iTAG, an approach to automatically perform the task of Intent Annotation. In Section V, we present the results from an exploratory study: attempting to automatically generate intent tags in a simplified setting, i.e. 44 political speeches given by Barack Obama and John McCain during the 2008 US presidential election campaigns. We evaluate our approach in Section VI, and conclude our work with a discussion of limitations and results in Sections VII and VIII.

The overall contribution of this paper is twofold:

- First, we discuss a novel and interesting dimension of tagging, and thereby expand the knowledge of tagging dimensions identified in the literature.
- Second, we present a prototypical method (iTAG) for automatically annotating textual resources with human intentions, and provide detailed evaluation results from a human subject study.

Our work thereby extends the repertoire of existing automatic tag generation techniques, and expands the knowledge that can be inferred from textual resources.

[1] A tag cloud is a non-hierarchical presentation of linked terms [12], often described as a visualization of word frequencies as well [21].

II. RELATED WORK

Two fields of related research are relevant: Studies of Folksonomies and work on Automatic Tag Generation.

A. Studies of Folksonomies

Bischoff et al. [2] analyze tagging behavior in four different datasets, i.e. flickr.com, del.icio.us, last.fm and web text anchors, and examined the kind of tags used, their distribution and their potential to improve search. Their work provides evidence for the empirical existence of different tagging dimensions, and shows the influence of the resource type, e.g. a text document versus an image, on tagging behavior. Golder and Huberman [7] examine structure and dynamical aspects of collaborative tagging systems – in particular in the context of del.icio.us. They introduce seven tag categories such as "Identifying what it is" or "Task Organizing". In another work, Heckner et al. [8] study four different social tagging platforms such as Flickr, del.icio.us, Youtube and Connotea and explore different types of behavior for different kinds of digital media. They also raise the question about the users' intent when annotating resources with tags. Heymann and Garcia-Molina [10] generate a navigable hierarchical taxonomy of tags from data of del.icio.us and CiteULike by evaluating the closeness centrality of the tags in analyzed networks. Such taxonomies can aid users in navigating folksonomies and help them to get a better overview of the tags already used in the system. In "Tags are not metadata, but just more content", Berendt and Hanser [1] suggest that tagging has the potential to add important information about resources that can be difficult to acquire from the resources themselves. Suchanek et al. [19] study the impact tag suggestions exert on the user. Their results indicate that the tag suggestion algorithm influences users' tagging behavior.

B. Techniques for Automatic Tag Generation

In the context of automatic tag generation, Chirita et al. [3] present *P-Tag*, a prototypical implementation of a personalized tag generation system. *P-Tag* attempts to automatically generate tags for visited web sites based on their content as well as on documents the user has on her computer to personalize the tag generation process. A paper by Sood et al. [17] presents *TagAssist*, a tool developed to support the process of tagging blog posts. For this purpose, tags of similar blog entries are aggregated and recommended to the user. Song et al. [16] introduce an automated framework to recommend tags for a new document which is added to a tagging system. They use graph clustering on two bipartite (tag – document) graphs to group together documents and tags and provide a ranking algorithm to propose tags for a new document to the user. Their tag recommendation technique labeled 88% of the test documents correctly.

III. INTENT ANNOTATION

Existing automatic tag suggestion approaches largely focus on annotating a document according to its predominant subject matter (what a resource is about, e.g. “sports” or “politics”). In this work, we aim to annotate resources according to the intentions described within them (what goals a resource is about, e.g. “Achieve Happiness” or “Maintain Good Health”). This type of annotation can be expected to introduce a new and interesting perspective on textual resources. *Intent Annotation* thereby represents an orthogonal view on topic annotation by attempting to answer *which intentions and human goals are referenced in a given textual resource*. Thereby, intent annotations deal with future states of affairs that some agent wants to achieve (goals), as opposed to topic, sentiment, or opinion tags where typically a present state is approximated. In addition, goals are frequently represented by compound tokens consisting of at least one verb and one or more other tokens (“looking young” as opposed to “youth”).

However, when examining a sample of web documents we observed that people rarely state their intentions explicitly in text, which makes the task of Intent Annotation an especially challenging endeavor. As an example, consider the human religious intention to ”Achieve Salvation“ (taken from [5]). Although this is an activity pursued by many, it is extremely rare to find someone who explicitly states her intent to accomplish this goal. However, people are quite prolific in writing about actions and activities they participate in on a daily basis, such as “*adhere to Jewish law*” or “*convert to Christianity*”, which can be assumed to indirectly contribute to “Achieve Salvation”.

In this work, we explore the use of such *indicative actions* as a proxy for inferring human intentions from textual resources. Intent Annotation can be understood as the problem of identifying a set of adequate intent annotations for each and every action indicative of intent in a given textual resource. More formally: Let $A = \{a_1, a_2 \dots a_n\}$ be a set of intent annotations and R be a domain of resources. Each document r_i comprises a sequence of sentences $S = \{s_1, s_2 \dots s_{|S|}\}$. The task of Intent Annotation is to approximate the unknown function $f\colon S \times A \rightarrow$ {True, False}, assigning the sentences a number of intent annotations ranging from 0 *to* $|A|$ (multi-label assignment). To give an example: A sentence “*I want to take care of my skin*” might be labeled with the intent annotation “Looking Young” as opposed to topic tags such as “beauty” or “skin”.

IV. AUTOMATIC INTENT ANNOTATION WITH iTAG

There are a number of alternative datasets that could be used as a basis for Intent Annotation, such as goals acquired from resources themselves, goals acquired from other resources (such as 43things.com, Search Query Logs, etc) or goals modeled in theoretical frameworks of human intent. In this work, we decided to base our iTAG automated annotation approach on the latter - an existing socio-psychological taxonomy of 135 categories of human intent [5]. This has two advantages: First, the theoretical framework was compiled by psychologists, and can be considered to be exhaustive to a certain extent by covering a broad range of different aspects of human intent. Second, the limited set of intent categories facilitates evaluation of our approach by transforming the large set of potential human goal instances into a manageable number of intent categories.

The iTAG approach presented in this paper consists of two building blocks: In a first step, we use the Web – accessed by Yahoo!’s BOSS search API – as a resource to build up a knowledge base that maps indicative actions to 135 intent categories. In a second step, we scan a given textual resource for indicative actions and assign corresponding intent categories.

A. Enriching a Taxonomy of Human Goals

We employed the social-psychological theoretical framework [5] that organizes high-level goals of people into 135 categories of human intent including ”A good marriage”, ”Getting an education“ and ”Taking care of family”. A useful property of taxonomies in general is that categories are hierarchically grouped into high-level categories, in our case top level categories such as ‘Family’, ‘Religion’ and ‘Money’ (not depicted in Figure 2). While we do not make use of hierarchical information in the current version of iTAG, using it in future work could help adding explicit relations between intent annotations.

Abbreviation	Full label
Achieving salvation	Achieving salvation
Arts	Appreciating the arts
Aspirations	Achieving my aspirations
Attracting sexually	Being able to attract, please, sexually excite a sexual partner
Avoiding failure	Avoiding failure
Avoiding guilt	Avoiding feelings of guilt
Avoiding rejection	Avoiding rejection by others
Avoiding stress	Avoiding stress
Being able to fantasize	Being able to fantasize, imagine
Being affectionate	Being affectionate toward others
Being ambitious	Being ambitious, hard-working
Being better than others	Being better than others, beating others

Figure 2 displays an excerpt of Chulef’s taxonomy of human goals [5]. The left part lists the first 12 intent categories and the right part provides additional information to each category.

While the taxonomy of human goals provides abbreviations and full labels for each intent category, further category descriptions are not available. In order to semantically enrich these category descriptions, we attempted to find corresponding descriptive phrases for each category. To give an example: Descriptive phrases for the category “Achieve Salvation” included “*to reach spiritual enlightenment*” or “*to get into heaven*”. The manual process of enriching the taxonomy with descriptive phrases was iterative. Together with Dr. Read, one of the co-authors of [5], we evaluated these mappings. During the evaluation phase, he helped us better understand intent category distinctions.

B. Constructing the Knowledge Base

We sought to generate a large knowledge base consisting of actions that indicate relevance for one of 135 categories. We attempted to acquire a large set of indicative actions by searching for sentences on the web (cf. [4]) that contained both (i) one of the descriptive phrases for the category, and (ii) an action-based causal relation. To achieve that, we constructed a series of query strings by concatenating each descriptive phrase with each of the following five causal relation phrases: "in order to", "for the purpose of", "essential for", "necessary for" and "critical for". Then, exact phrase searches were issued to the web using the Yahoo! BOSS API[2] for all constructed query strings. The textual content of the first 500 result pages was retrieved, parsed and sentence delimited. Sentences that contained query phrases were stored in our knowledge base, which was implemented via an Apache Lucene[3] index. Table 1 shows sample phrase queries and retrieved sentences with the respective indicative actions underlined.

Table 1 shows exemplary query strings for the category "Looking Young" and retrieved sentences containing indicative actions.

Query string causal relation + *descr. Phrase*	**Retrieved Sentences (Yahoo)** indicative actions
"in order to *look young*"	In order to *look young* and beautiful, you need to take care of your skin.
"for the purpose of *looking young*"	While we know that fitness is one of the keys to remaining healthy, *we* also exercise for the purpose of *looking young and sexy*.
"in order to *look youthful*"	It was in the context of people drinking a lot of water in order to *look youthful*.
"in order to *avoid wrinkles*"	You need to moisturize inside and out, in order to *avoid wrinkles*.

C. Matching Sentences to Intent Categories

To automatically generate intent annotations for a given textual resource, we first segment the document into a set of sentences for subsequent analysis. Then, each sentence in the document is issued as a query to the knowledge base using Lucene's default similarity measure. This allows identifying the most similar sentence in our knowledge base containing indicative actions. We required a similarity greater than 0.5 (1.0 equals an exact match) as a quality criterion of the retrieved sentences. Then the intent category associated with the knowledge base entry is assigned as the intent annotation in a 'Winner takes it all' approach. Intent annotations for entire documents are produced by aggregating intent annotations of all sentences.

[2] http://developer.yahoo.com/search/boss/

[3] http://lucene.apache.org/

V. RESULTS

To gauge the prospects of intent annotation, we applied our approach to a limited set of textual documents that we suspected to be particularly amenable for our purposes. Due to the exploratory nature of our research, we decided to use political speeches over other textual resources such as blog posts, because (i) political speeches typically have a clear focus on discussing, conveying or achieving goals (ii) transcripts of political speeches are less affected by noise compared to other resources, and (iii) political speeches can be expected to contain a broad variety of intentions. These factors *facilitate evaluation* and make political speeches particularly suitable to explore the prospects of intent annotation in a *simplified setting*. In the future, we are interested in applying our approach to more challenging settings such as search query logs, blog posts, twitter feeds or discussion boards, where additional challenges such as lack of focus, noise and other problems would have to be addressed.

We retrieved and preprocessed the textual resources of 44 transcripts of political speeches given in April and June 2008 by the two leading American presidential candidates, John McCain and Barack Obama. After data cleansing and sentence delimitation, every sentence was treated as a query for the knowledge base.

A. Intent Annotation

Figure 4 depicts selected results of applying iTAG to speeches given by Obama. The matrix shows 21 speeches and their relation to 135 categories of human intent. Each cell contains a weight describing the relative importance of a given goal category for a particular speech. From this figure, we can see that certain intent categories dominate throughout all speeches analyzed in our study such as the intent category "Helping Others" or "Charity", while other categories exhibit temporal bursts, for example "Being Better Than Others". The data can be analyzed from a number of perspectives. In Figure 3 for example, intent categories for Barack Obama's and John McCain's speeches are contrasted.

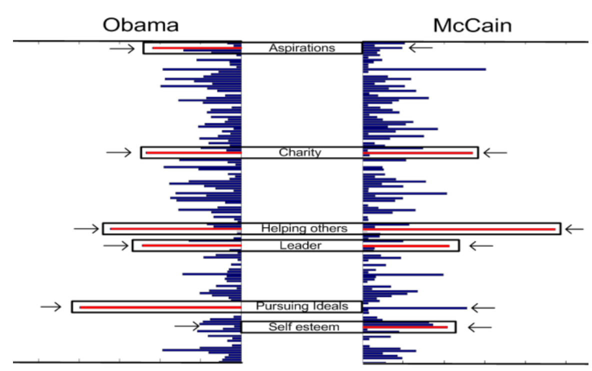

Figure 3 compares Intent Categories for Obama's and McCain's speeches. Results are averaged over 44 (21+23) speeches (April and June 2008). Predominant categories such as "Charity" are highlighted.

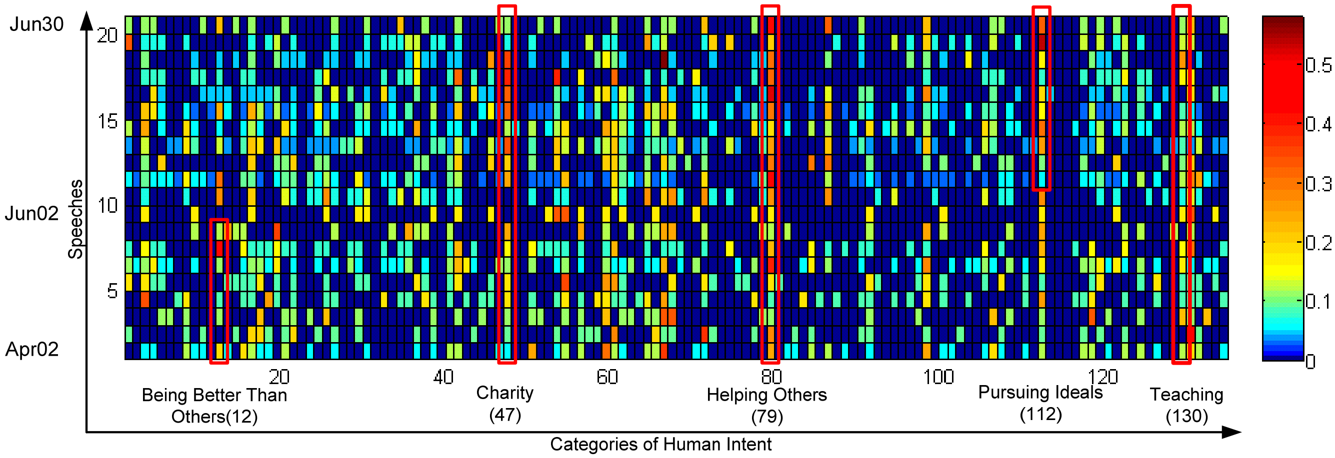

Figure 4 provides an overview of intent annotations for 21 speeches given by Barack Obama in April and June 2008. Selected categories which are predominant over a certain period of time are highlighted.

At a first glance, similarities and differences between the two candidates can easily be recognized, providing some sort of intentional summary of the speech contents. Both candidates conveyed messages to their audience that were often assigned to high-level intent categories such as "Leader", "Helping Others" and "Charity". Figure 3 also reveals intent categories that are stronger associated with one of the two candidates: categories such as "Self Esteem" have a higher weight for McCain's speeches whereas Obama's speeches seem to emphasize other categories such as "Pursuing Ideals" and "Aspirations".

The mapping of sentences to categories of human intent can now be used to produce intent annotations for each of the 44 speeches. The iTAG automatic Intent Annotation approach yields a ranked list of intent annotations based on the 25 most dominant intent categories identified for a given textual resource. Figure 5 and Figure 6 present tag clouds of intent annotations for speeches given by Obama and McCain. The text size of intent tags is based on the weight of annotations assigned to Barack Obama's and John McCain's speeches. Text size in our clouds scales linear; to visualize the clouds we used existing online services[4]. In both cases the top 25 tags were retained.

While the tag clouds depicted in Figure 5 and Figure 6 aggregate intent annotations for a number of speeches given by the candidates, iTAG could be applied on an individual speech and/or passage level as well, assuming the presence of a sufficient number of sentences containing indicative actions.

Figure 5 shows a tag cloud of Intent Annotations for 21 speeches given by Barack Obama.

The two tag clouds presented in Figure 5 and Figure 6 reveal further interesting differences between the goals pursued by the two presidential candidates. While McCain's most dominant goals are "Helping Others" and "Being better than others", "Pursuing ideals" and "Helping Others" represent the highest-weighted annotations for Obama, according to iTAG. This is an interesting, yet anecdotal, result concurring with a popular media characterization of Obama's political motivations as driven by and aspiring to ideals.

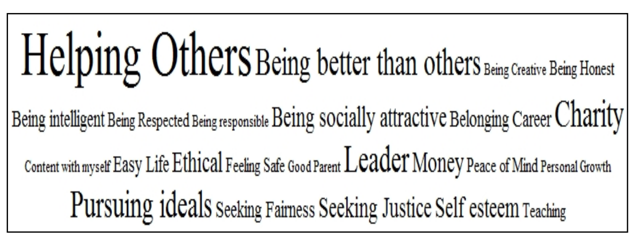

Figure 6 shows a tag cloud of Intent Annotations for 23 speeches given by John McCain.

B. Evolution of Intent Annotations

Because we have temporal information about the date of the speeches, a number of interesting additional analyses can be conducted. For example, Figure 7 illustrates the temporal evolution of intent annotations over 21 speeches given by Barack Obama in April and June 2008.

---

[4] http://www.tocloud.com/javascript_cloud_generator.html

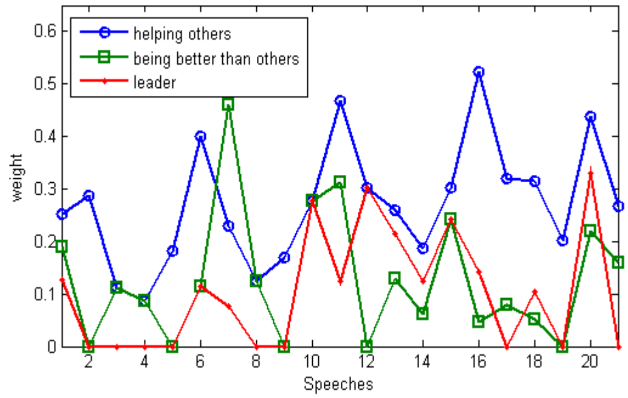

Figure 7 illustrates temporal evolution chart of three selected Intent Annotations "Helping Others", "Being Better Than Others" and "Leader" over 21 speeches given by B. Obama.

Several observations can be made when focusing the comparison on a few selected annotations. For example, the chart in Figure 7 shows that intent annotations such as "Helping Others" are prominent over the entire period that was observed. Peaks where individual intent annotations dominate can be easily detected, such as "Being better than others" in speech No. 7 or "Leader" in speech No. 20. It is conceivable that applying this type of analysis to other resources, such as search query logs, blog posts or discussion forums, could open up new opportunities to interlink textual resources on the web or to monitor social media activities.

VI. EVALUATION

A. Usefulness of Knowledge Base

On a general level, the usefulness of a knowledge base for the purpose of intent annotation can suffer from a single or a combination of the following issues: (i) the knowledge base entry does not contain an indicative action, (ii) the entry contains an action but it is unrelated to the corresponding intent category and (iii) the entries for a given category only represent a minor fraction of possible actions. Combined, these factors have the potential to introduce noise and bias to the knowledge base. In the following, we aim to estimate the usefulness of the iTAG knowledge base by investigating qualitative and quantitative aspects.

The minimum number of knowledge base entries per category was 12 (Category: "Firm Values"), the maximum number was 4,497 (Category: "Helping Others") and the average number was 752. The final number of sentences in the knowledge base totaled 101,490. The distribution of knowledge base entries is skewed as depicted in Figure 8 yet only a minor fraction of categories received less than 100 entries.

In order to evaluate the quality of knowledge base entries, we drew a random sample of 674 entries from the knowledge base. The sample was judged by a linguistics undergraduate student with regard to 1) whether the entry contains indicative actions and 2) whether the entry is relevant for the corresponding category.

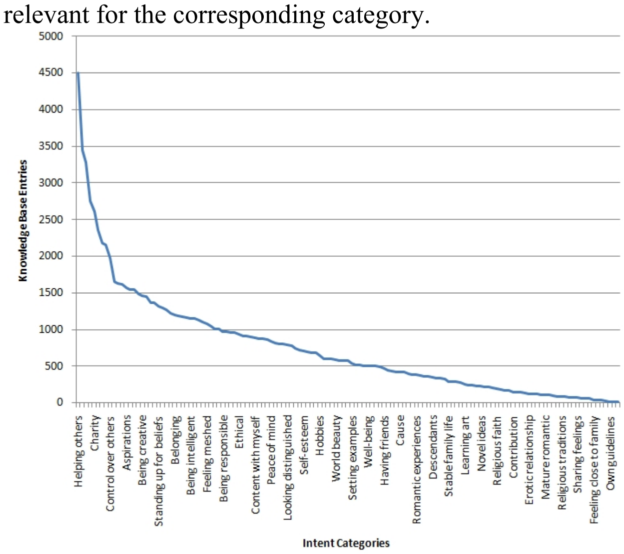

Figure 8 shows the distribution of knowledge base entries per category.

57% of the entries in the sample contained actions indicative of the corresponding intent category, which says that while there is a certain level of false positives, the majority of entries are useful. To evaluate relevance of knowledge base entries, we conducted comparative analyses of different causal relations. Table 2 shows success rates for every causal relation where success rate is defined as #correct entries divided by #all entries regarding a particular causal relation. Two relations, i.e. "in order to" and "for the purpose of", exhibit success rates beyond 50% suggesting better quantities and higher-quality knowledge better entries than others, such as "essential for" and "necessary for". For the results reported in this paper, only the causal relations "in order to" and "for the purpose of" were used, information acquired through other relations was discarded as a result of this evaluation. We are well aware that our choice of causal relations does not cover all potential intent – action pairs. By restricting ourselves, we certainly miss many pairs, yet, we do not aim to achieve an optimal coverage of intent – action pairs for every category but a reasonably sufficient one.

Table 2 illustrates the quality of the five used causal relations. Only those exhibiting a success rate beyond 50% were taken into account for further processing steps.

	in order to	essential for	necessary for	critical for	for the purpose of
Success Rate	59.2 %	32%	35.5%	16.7%	59.8%

In addition to this evaluation, we wanted to investigate people's agreement on sentences that contain indicative actions. Using Cohen's Kappa coefficient [6], we obtained κ =0.79. This indicates that human annotators can largely agree on what constitutes suitable entries in our knowledge base.

B. Automatic vs. Manual Intent Annotation

To evaluate the quality of automatic intent annotations, we compared annotations produced by iTAG with the annotations produced in a human subject study where all three annotators were Computer Science graduate students. We had two human subjects annotate Obama's speeches and assign intent annotations to sentences that the subject believed would contain indicative actions. The annotators judged 3,722 sentences from 21 speeches and agreed upon 3,382 sentences to either assign no or at least one category to the same sentence. The corresponding Kappa κ =0.82 reflects useful agreement amongst the raters.

Table 3 compares the top 25 annotations produced by iTAG and human annotators for Obama's speeches. Weights represent normalized frequency values. Highlighted entries represent entries that were assigned by both iTag and human annotators.

iTAG Annotation 21 Speeches by Barack Obama		Rank	Manual Annotation 21 Speeches by Barack Obama	
Goal Category	Weight		Goal Category	Weight
Pursuing ideals	0.1991	1	Helping others	0.1798
Helping others	0.1615	2	Contribution	0.0944
Leader	0.1223	3	Difficult things	0.0831
Charity	0.1177	4	Bills	0.0742
Aspirations	0.1094	5	Job	0.0607
Being free	0.0993	6	Seeking equality	0.0494
Teaching	0.0968	7	Charity	0.0449
Being better than	0.0965	8	Education	0.0404
Control over others	0.0956	9	Feeling safe	0.0404
Being creative	0.0940	10	Being better than	0.0382
Education	0.0886	11	Seeking fairness	0.0382
Exercising	0.0874	12	Being responsible	0.0270
Ethical	0.0807	13	Being ambitious	0.0247
Exploring	0.0797	14	Money	0.0247
Feeling safe	0.0771	15	Being innovative	0.0202
Being likeable	0.0771	16	Control over others	0.0157
Content with myself	0.0762	17	Seeking justice	0.0157
Money	0.0721	18	Avoiding failure	0.0157
Attracting sexually	0.0709	19	Overcoming failure	0.0157
Knowing many	0.0645	20	Teaching	0.0112
Easy life	0.0644	21	Providing family	0.0112
Being curious	0.0576	22	Own guidelines	0.0112
Avoiding stress	0.0550	23	Close children	0.0090
Sexual experiences	0.0540	24	Leader	0.0007
Being self-sufficient	0.0530	25	Being free	0.0045

In case of McCain's speeches, we had a single human subject annotate McCain's 23 speeches, altogether 2,677 sentences. We used the manual intent annotations to produce a ranking of intent categories for each candidate. In case of Obama's speeches, we took the union of annotations produced by the two human subjects to mitigate data sparsity.

Table 3 and Table 4 present the most frequent annotations produced by iTAG and the human annotators based on the aggregation of 21 speeches by Barack Obama and 23 speeches by John McCain. Out of the top 25 intent categories produces by iTAG, the annotations produced by human annotators agreed with the automated iTAG approach in 10 cases (40%). Agreement for McCain's speeches was similar, with 11 (44%) tags shared by iTAG and the human annotation ranking. Highlighted entries in Table 3 and Table 4 represent entries that were assigned by both iTAG and human annotators.

Table 4 compares the top 25 annotations produced by iTAG and human annotators for McCain's speeches. Weights represent normalized frequency values. Highlighted entries represent entries that were assigned by both iTag and human annotator.

iTAG Annotation 23 Speeches by John McCain		Rank	Manual Annotation 23 Speeches by John McCain	
Goal Category	Weight		Goal Category	Weight
Helping others	0.2368	1	Avoiding failure	0.0958
Being better than	0.1513	2	Aspirations	0.0949
Charity	0.1350	3	Standing up for	0.0873
Pursuing ideals	0.1278	4	Helping others	0.0863
Leader	0.1058	5	Being respected	0.0852
Self esteem	0.1039	6	Pursuing ideals	0.0586
Ethical	0.1030	7	Being recognized	0.0543
Money	0.0990	8	Persuading others	0.0383
Being socially	0.0919	9	Being responsible	0.0362
Seeking justice	0.0862	10	Overcoming failure	0.0319
Seeking fairness	0.0811	11	Novel ideas	0.0309
Being intelligent	0.0805	12	Own guidelines	0.0277
Easy life	0.0773	13	Leader	0.0266
Belonging	0.0747	14	Support from others	0.0266
Career	0.0738	15	Being better than	0.0191
Peace of mind	0.0673	16	Control over others	0.0181
Being honest	0.0653	17	Teaching	0.0170
Teaching	0.0651	18	Others' trust	0.0170
Feeling safe	0.0643	19	Seeking fairness	0.0170
Being respected	0.0616	20	Being honest	0.0160
Being creative	0.0590	21	Seeking justice	0.0150
Good parent	0.0567	22	Freedom of choice	0.0128
Personal growth	0.0543	23	Career	0.0128
Content with	0.0529	24	Seeking equality	0.0110
Being responsible	0.0525	25	Taking care of family	0.0110

In order to gauge the quality of intent annotations produced by iTAG, we used the top 25 manual annotations as relevant annotations (right columns in Table 3 and Table 4) and judged the remaining manual annotations to be irrelevant. Using the manual annotations as our "ground truth", Figure 9 and Figure 10 show the performance of iTAG annotations in comparison to a simple baseline approach. The baseline approach ranks intent categories in a random manner.

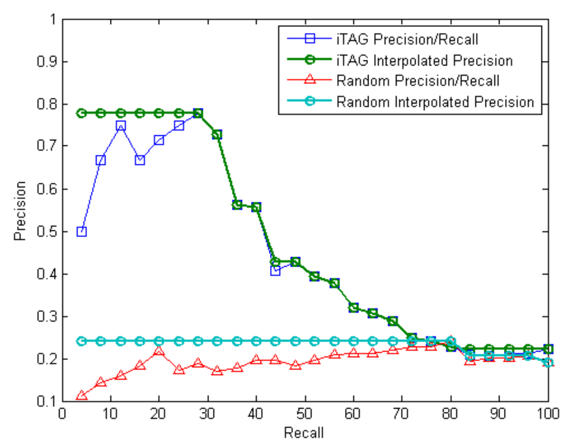

Figure 9 compares the iTAG vs. random approaches for Obama's speeches in terms of precision and recall.

Figure 11 provides a visual comparison between Intent- and Traditional Tag Clouds based on all speeches given by Barack Obama and John McCain. The figure should illustrate that there is no rivalry between intent and traditional intent tags, yet they rather complement each other by providing two different perspectives onto political speeches.

Both Figure 9 and Figure 10 show that our iTAG approach outperforms the simple baseline approach for recall levels of up to 70%. The results illustrate that for up to 40% recall (10 relevant annotations), the iTAG approach achieves a precision of 50% and above. While there is room for improvement, the results demonstrate the principle feasibility of automatically annotating textual resources with human intent and represent a first step towards more sophisticated approaches.

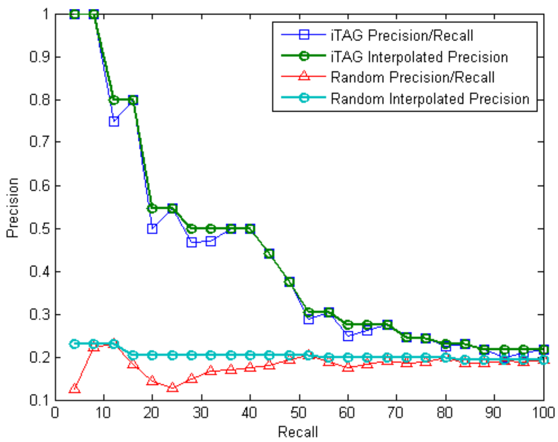

Figure 10 compares the iTAG vs. random approaches for McCain's speeches in terms of precision and recall.

C. Intent vs. Content Annotations

In order to visually illustrate the differences between intent and content annotations, we produced different tag clouds – Intent Tag Clouds and Traditional Tag Clouds – from the same data. Intent tag clouds were produced by iTAG, while "traditional" tag clouds were produced by counting word occurrences in the text, and eliminating words based on a list of stop words [23]. Figure 11 illustrates an excerpt of the tag clouds produced. On the right hand side, traditional tag clouds of McCain's and Obama's speeches are presented, while on the right hand side of Figure 11, intent tag clouds show a different perspective on the same data. We can see that while the traditional tag clouds provide a rough overview of the vocabulary used by the two candidates, intent tag clouds highlight the goal categories that are most important to them.

VII. DISCUSSION AND LIMITATIONS

In the following, we discuss selected threats to validity to our work:

A. Usefulness of the Knowledge Base:

While our knowledge base was helpful to produce intent annotations that achieve a useful level of agreement with human annotators, it suffers from (i) a skewed distribution of #entries per category, (ii) a certain amount of false-positive indicative actions, (iii) and noise. We addressed some of these concerns in our work (e.g. by eliminating causal relations that tend to produce false positives), but there are several opportunities to build on and improve our results in future work.

To study whether the skewed distribution of entries in the knowledge base biases the automatic intent annotation task, we conducted additional analyses. We calculated Spearman's rank correlation between the ranked list of knowledge base categories (where the category with the highest number of entries ranks first) and the categories produced by iTag for the speeches by Obama and McCain. The results of this calculation reveal that there is a weak correlation between the ranked intent categories and iTag's annotations, i.e. Obama = 0.38 and McCain = 0.42. This corroborates that our current approach is – to a certain extent – biased towards the number of entries per category in the knowledge base.

B. Quality of Automatic Intent Annotation:

The process of automatically generating intent annotations faces a number of challenges, which we'd like to explain by using an illustrative example: Consider the search query: "in order to *age well*", which corresponds to the intent category "Looking Young". Among other results, this query could produce the following problematic search result:

"Cork has been used for over 400 years, and many winemakers today still believe that in order to *age well*, wine <u>needs gradual exposure to oxygen</u>."

Such problems cause sentences being misclassified, and negatively influence results. However, because iTAG is based on aggregating evidence and taking a "winner takes all approach", it is tolerant against occasional misclassification of sentences, and our evaluation of knowledge base entries revealed that a majority of indicative actions represent suitable proxies for the automated intent annotation task. However, an option to reduce this problem in future work could be to employ parsing to alleviate the semantic problem (cf. [21]) and/or using machine learning techniques to distinguish between sentences that should be assigned intent categories and those that should not. While our approach outperforms a random ranking of annotations, further comparisons with other approaches need to be conducted in future research.

C. Matching Sentences to Intent Categories

In our current approach, we employ Apache Lucene's retrieval functionality to query the knowledge base and obtain most similar knowledge base entries. By default, the similarity calculation is based on a bag-of-words approach that neglects the word order. Future work might explore the usage of more sophisticated similarity measures or explore the use of n-grams that could provide extra lexical information.

VIII. CONCLUSIONS

Intent Annotations add interesting information to textual resources, which is difficult to extract from the resource itself. In the past, automatic tag generation approaches demonstrated their usefulness in a broad range of different applications, including tag suggestion, resource clustering, resource enrichment or tag-based navigation. Our work adds a novel dimension to the set of tag dimensions identified in the literature. The prototypical iTAG method demonstrates the principle feasibility of automated intent annotation in a simplified setting, i.e. 44 political speeches, and thereby extends the repertoire of existing automatic tag generation techniques. In this sense, our work contributes to expanding the knowledge that can be inferred from textual resources and thereby it has the potential to open up new perspectives in the area of text understanding as well. Although our approach commits to a particular categorization schema for human goals (the socio-psychological theoretical framework consisting of 135 goal categories [5]), the general problem of Intent Annotation is agnostic with regard to the source of annotations, and other sources of intent annotations are conceivable.

The iTAG approach presented in this paper could help to open up a new intentional dimension to navigating and browsing textual resources on the web. While we have shown that intent annotations produced by iTAG achieve useful agreement with human annotators, more research is necessary to further improve accuracy of annotations.

To enable playful experimentation with the iTAG approach, we make a user interface available via http://webdev.know-center.tugraz.at:8080/intenttagcloud/. The web interface takes arbitrary textual contents as input and outputs corresponding intent tag clouds.

ACKNOWLEDGMENT

Thanks to E. Tavano, J. Liegl, H. Stern and D. Zibold for participating in evaluation tasks and to S. Read for assistance regarding the human goal taxonomy. Special thanks to Andrew S. Gordon and Reid Swanson at the University of Southern California's Institute for Creative Technologies for their assistance in completing this research. This work is funded by the FWF Austrian Science Fund Grant P20269 TransAgere. The Know-Center is funded within the Austrian COMET Program under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

REFERENCES

[1] Berendt B. and Hanser C.: Tags are not metadata, but "just more content" - to some people. In Proceedings of the International Conference on Weblogs and Social Media (ICWSM'07), Boulder, CO, USA, 2007.

[2] Bischoff K., Firan C. S., Nejdl W. and Paiu R.: Can all tags be used for search? In Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM'08), ACM, New York, NY, USA, 2008.

[3] Chirita P. A., Costache S., Nejdl W. and Handschuh S.: P-TAG: Large scale automatic generation of personalized annotation tags for the web. In Proceedings of the 16th International Conference on World Wide Web (WWW'07), ACM, NY, USA, 2007.

[4] Cimiano P. and Staab S.: 'Learning by googling'. In SIGKDD Exploration Newsletter. 6(2), 2004.

[5] Chulef A. S., Read S. J. and Walsh D. A.: A hierarchical taxonomy of human goals. In Motivation and Emotion, 25 (3) 2001.

[6] Cohen J.: A coefficient of agreement for nominal scales. In Educational and Psychological Measurement 20(1), 37, 1960.

[7] Golder S. A. and Huberman B. A.: Usage patterns of collaborative tagging systems. In Journal of Information Science. 32(2), 2006.

[8] Heckner M., Neubauer T. and Wolff C.: Tree, funny, to_read, google: what are tags supposed to achieve? A comparative analysis of user keywords for different digital resource types. In Proceedings of the 2008 ACM Workshop on Search in Social Media (SSM'08), 2008.

[9] Heymann P., Koutrika G. and Garcia-Molina H.: Can social bookmarking improve web search? In Proceedings of the International Conference on Web Search and Web Data Mining (WSDM'08), ACM, NY, USA, 2008.

[10] Heymann P. and Garcia-Molina H.: Collaborative creation of communal hierarchical taxonomies in social tagging systems, Technical Report, Computer Science Department, 2006.

[11] Hotho A., Jäschke R., Schmitz C. and Stumme G.: BibSonomy: A social bookmark and publication sharing system. In Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures, 2006.

[12] Kaser O. and Lemire D.: Tag-cloud drawing: Algorithms for cloud visualization, May 2007.

[13] Lambiotte R. and Ausloos M.: Collaborative tagging as a tripartite network. Arxiv preprint cs.DS/0512090 v2, 2005.

[14] Mika, P.: Ontologies are us: A unified model of social networks and semantics. In Proceedings of the International Semantic Web Conference (ISWC'05), 2005.

[15] Pang B. and Lee L.: Opinion mining and sentiment analysis. In Foundations and Trends® in Information Retrieval 2(1-2), 2008.

[16] Song Y., Zhuang Z., Li H., Zhao Q., Li J., Lee W. and Giles C. L.: Real-time automatic tag recommendation. In Proceedings of the 31st Annual International Conference on Research and Development in Information Retrieval (SIGIR '08), ACM, NY, USA, 2008.

[17] Sood S., Owsley S., Hammond K. and Birnbaum L.: TagAssist: Automatic Tag Suggestion for Blog Posts. In Proceedings of the International Conference on Weblogs and Social Media (ICWSM'07), 2007.

[18] Strohmaier M.: Purpose Tagging - Capturing User Intent to Assist Goal-Oriented Social Search. In Workshop on Search in Social Media (SSM'08), in conjunction with CIKM'08, Napa Valley, USA, 2008.

[19] Suchanek F. M., Vojnovic M. and Gunawardena D.: Social tags: meaning and suggestions. In Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM'08), New York, NY, USA, 2008.

[20] Turney P. D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, Morristown, NJ, USA, pp. 417—424, 2002.

[21] Wiebe J. and Mihalcea R.: Word sense and subjectivity. In Proceedings of the 21st International Conference on Computational Linguistics (ACL'06), Association for Computational Linguistics, Morristown, NJ, USA, 2006.

[22] Viegas F.B., Wattenberg M., van Ham F., Kriss J., and Mckeon M.: Many Eyes: a site for visualization at internet scale. Transactions on Visualization and Computer Graphics, 13(6), 2007.

[23] http://www.dcs.gla.ac.uk/idom/ir_resources/linguistic_utils /stop_words, last accessed February 15th 2009.

Mark Kröll is a PhD student at the Knowledge Management Institute at Graz University of Technology since 2008. His current research focuses on automatically acquiring knowledge about common human goals from the web.

Christian Körner is a PhD student at the Knowledge Management Institute at Graz University of Technology since 2009. His current research focuses on studying the motivation behind tagging.

Dr. Markus Strohmaier is an Assistant Professor at the Knowledge Management Institute at Graz University of Technology since 2007. From 2006 to 2007, he was a Post-Doctoral Fellow at the Department of Computer Science at University of Toronto. His current research focuses on knowledge acquisition from the web.