Towards Automatically Annotating Textual Resources with Human Intentions

Markus Strohmaier, Mark Kröll and Christian Körner

Knowledge Management Institute, Graz University of Technology, A-8010 Graz, Austria

(markus.strohmaier@tugraz.at, mkroell@tugraz.at, christian.koerner@tugraz.at)

Problem Statement

How and to what extent is it feasible to automatically annotate textual resources with human intentions?

Motivation

-) Intent plays a fundamental role in user interactions on the web, including interpreting and understanding resources.
-) We introduce Human Intentions as new and interesting annotation dimension.
-) In contrast to traditional annotations such as topic or quality, intent annotations focus on future states of affairs.
-) 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".
-) Potential benefits: (i) to quickly grasp the main aspirations implicitly addressed by resources or (ii) to enable goal-oriented navigation of resources, such as blogs, on the web.

Intent Annotation

-) Intent Annotation can be understood as the problem of automatically identifying a set of adequate intent annotations for textual resources.
-) Intent Annotation approximates the unknown function $f: S \times C^I \rightarrow \{True, False\}$, where $C^I = \{c^I_1, c^I_2 \ldots c^I_n\}$ is the set of predefined intent categories, D is a domain of text documents and each document d_i consists of a sequence of sentences $S = \{s_1, s_2 \ldots s_{|S|}\}$.
-) Observation: people rarely state their intentions explicitly in text, which makes the task of Intent Annotation an especially challenging endeavor.
-) Example: consider the human religious intention to "Achieve Salvation" (taken from [1]). Although this is an activity pursued by many, it is extremely rare to find someone who states their plan on how to accomplish this goal. However, people are quite prolific in writing about the actions and activities they participate in, such as "convert to Christianity", which can be assumed to indirectly contribute to "Achieve Salvation".
-) We explore indicative actions as a proxy for inferring intentions from textual resources.

Approach: iTAG Automatic Prototype

<u>1.) Enriching a Taxonomy of Human Goals:</u> We employed the social-psychological theoretical framework [1] 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". We compiled a set of 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".

<u>2.) Constructing the Knowledge Base:</u> We attempted to acquire indicative actions by searching for sentences on the web (cf. [2]). We constructed a series of query strings by concatenating each descriptive phrase with causal relation phrases, e.g. "in order to". Then, exact phrase searches were issued to the web using Yahoo! BOSS. We identified ~169.000 sentences that contained the query phrases along with indicative actions. These sentences were stored in our knowledge base, an Apache Lucene index.

<u>3.) Matching Sentences to Intent Categories:</u> We first segment the document into a set of sentences. Each sentence in the document is issued as a query to the knowledge base to identify the most similar sentence in our knowledge base.
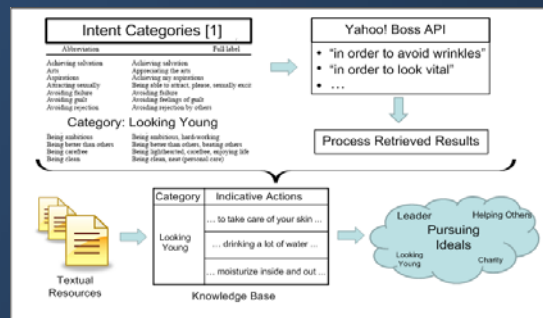

Figure 1 visualizes the process of automatically annotating textual resources with human intentions.

Results & Evaluation: Automatically Annotating B. Obama's and J. McCain's US Presidential Election Speeches

We applied our approach to the textual resources of 44 transcripts of political speeches given by the two American presidential candidates in 2008. Figure 2 provides a visual comparison between traditional and intent tag clouds. To evaluate the quality of automatic intent annotations, we compared annotations produced by iTAG with the annotations produced in a human subject study. Moreover, a precision/recall curve was calculated to contrast iTag against a random approach (see Figure 3).

-) out of the top 25 intent categories produced by iTAG, 11 categories (44%) are shared with the human annotations (ground truth)
-) our iTAG prototype outperforms a simple baseline approach for recall levels of up to 70%
-) our iTAG prototype achieves a precision of 50% and above for 40% recall (10 relevant annotations)

Figure 2: Visual Comparison between Intent- and Traditional Tag Clouds based on Transcripts of Political Speeches Given by B. Obama and J. McCain in 2008.
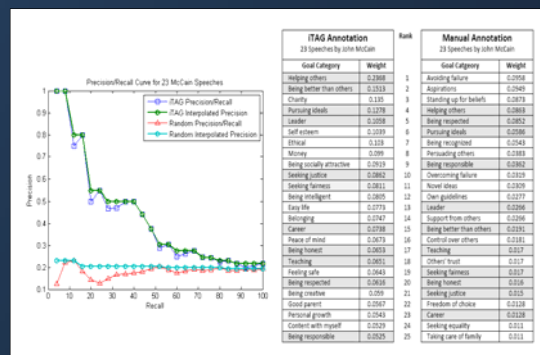

Figure 3: Comparison of the top 25 Intent Annotations produced by iTAG and by the human raters for 23 speeches of J. McCain. Highlighted entries were assigned by both iTAG and human annotators. The precision/recall curve on the left hand side evaluates iTag against a random approach.

Conclusions

This work
-) adds a novel dimension to the set of tag dimensions: Human Intentions
-) expands the knowledge that can be inferred from textual resources
-) demonstrates the principle feasibility of automated intent annotation
-) thereby extends the repertoire of existing automatic tag generation techniques [3]

Further Questions

1) How can we employ this new dimension to aid navigating and to browsing textual resources on the web?
2) To what extent is our approach agnostic to other textual corpora such as weblogs?

[1] Chulef, A. S., Read, S. J. and Walsh, D. A.: A Hierarchical Taxonomy of Human Goals. In Motivation and Emotion, 25 (3), 191--232, 2001.
[2] Cimiano, P. and Staab, S. : 'Learning by googling', SIGKDD Explor. Newsl. 6(2), 24—33, 2004.
[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 'WWW '07: Proceedings of the 16th international conference on World Wide Web', ACM, New York, USA, 2007.

Check out our online demonstrator:
http://webdev.know-center.tugraz.at:8080/awm09/index.jsp