Analyzing Human Intentions in Natural Language Text

Mark Kröll[1] and Markus Strohmaier[1,2]

Knowledge Management Institute, Graz University of Technology[1] and Know-Center[2], 8010 Graz, Austria

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

McCAIN vs. Obama'08 BarackObama.com

Problem Statement

How and to what extent is it possible to automatically analyze human intentions in textual content?

Motivation

A Novel Perspective:

-) existing text analysis techniques provide different perspectives on textual resources, e.g. topics, sentiment or opinion analysis [3]
  → Can we analyze textual content from an intentional perspective?

The Challenge:

-) people rarely state their intentions explicitly in text which makes analyzing human intentions in text a difficult endeavor

The Opportunity:

-) 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" (taken from [1])
  → Can we use these *indicative actions* as a proxy for inferring intentions?

Consequences:

-) inferring human intentions from textual content could contribute to
  (i) annotating resources with intentions referenced within them
  (ii) answering "Why" questions in Natural Language Text applications

Intent Analysis

Definition:

-) We define Intent Analysis as the problem of automatically identifying a set of adequate intent annotations for textual resources.
-) Intent Analysis 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 $S = \{s_1, s_2 \ldots s_{|S|}\}$.
-) yields an *intent profile vector* $ip \in R^n$, when applied to a text document, i.e. $f(d)$

Demarcation:

-) traditional analysis techniques such as topic, sentiment or opinion analysis [3]) aim to approximate a present (e.g. an emotional) state
-) in contrast, Intent Analysis deals with a different temporal focus, i.e. a *future state of affairs*, illustrating the difference between **Content vs. Intent**

Approach: Automatically Generating Intent Profiles

1.) Enriching a Taxonomy of Human Goals: 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*".

2.) Constructing the Knowledge Base: 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.

3.) Matching Sentences to Intent Categories: 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.

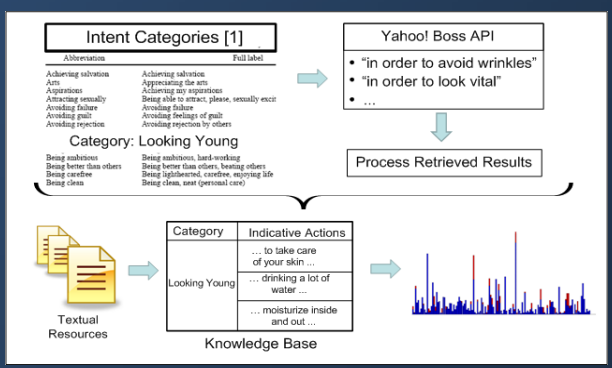

Figure 1: This figure visualizes the process of automatically analyzing textual resources and generating intent profiles from textual resources.

Results & Evaluation: Automatically Analyzing 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 the two candidates intent profiles. To evaluate the quality of automatic intent annotations, we compared notations produced in a human subject study.

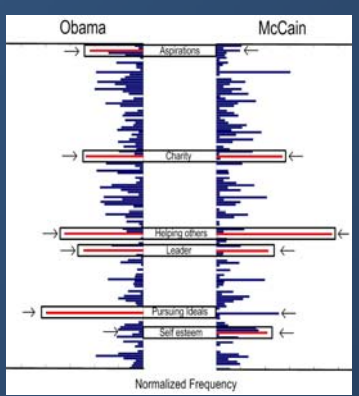

Figure 2: Intent profiles of Obama and McCain, generated from and averaged over 44 speeches (April and June 2008). It illustrates similarities and differences between the two candidates that can be easily recognized (intentional summary).

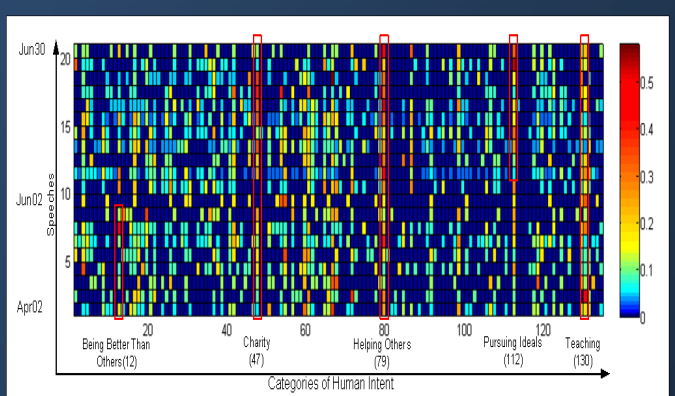

Figure 3: This visualization allows a comparison of intent profiles over time. The intent profiles were generated from 21 speeches given by Barack Obama in April and June 2008. It shows the temporal evolution of intent in the political speeches. Categories such as "Charity" and "Helping Others" are prominent over the whole period. Peaks such as the category "Pursuing Ideals" can be easily detected.

Automatic 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 others	0.1513	2	Aspirations	0.0949
Charity	0.135	3	Standing up for beliefs	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	0.103	7	Being recognized	0.0543
Money	0.099	8	Persuading others	0.0383
Being socially attractive	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 others	0.0191
Peace of mind	0.0673	16	Control over others	0.0181
Being honest	0.0653	17	Teaching	0.017
Teaching	0.0651	18	Others' trust	0.017
Feeling safe	0.0643	19	Seeking fairness	0.017
Being respected	0.0616	20	Being honest	0.016
Being creative	0.059	21	Seeking justice	0.015
Good parent	0.0567	22	Freedom of choice	0.0128
Personal growth	0.0543	23	Career	0.0128
Content with myself	0.0529	24	Seeking equality	0.011
Being responsible	0.0525	25	Taking care of family	0.011

Figure 4: Comparison of the top 25 Intent Categories (automatically and manually generated) for 23 speeches of J. McCain. The results illustrate that 11 categories (44%) are shared between automatic and human annotations.

Contributions

This work

-) adds a novel dimension to the repertoire of textual data analysis techniques
-) expands the knowledge that can be inferred from textual resources
-) demonstrates the possibility of automated intent analysis
-) indicates the potentials of Intent Analysis as a quick, visual evaluation of text from an intentional perspective

Further Questions

1) To what extent is intent analysis applicable to other text corpora such as weblogs?
2) To what extent does our approach contribute to the notion of social media monitoring?

Acknowledgements:
Thanks to E. Tavano, J. Liegl, H. Stern and D. Zibold for conducting the annotation tasks and to S. Read for assistance regarding the human goal taxonomy. Special thanks to A.S. Gordon and R. Swanson at the ICT, University of Southern California, for their assistance in completing this research.

[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] Pang, B. and Lee, L. Opinion mining and sentiment analysis.In Foundations and Trends® in Information Retrieval 2(1-2), 1--135, 2008.