

Acquiring Knowledge about Human Goals from Search Query Logs

Markus Strohmaier, Peter Prettenhofer and Mark Kröll

Motivation

- Knowledge about human goals has been found to be important
 - **Queries containing user goals:**
 - goal recognition for user actions (navigation)
 - the generation of action sequences that implement goals (planning) [Lieberman07]
 - how to lose weight
- How do we define a query containing a user goal?
 - turning blonde hair to dark brown
 - download pictures of angels

Related Research:

- **Understanding queries not containing user goals.** [Sergeev02, Rose and Levinson 04]
 - high level categorization of queries to improve retrieval on the web
 - dining room furniture
 - yet, we know little about the specific goal instances
 - you had a bad day
- **Goals and common sense** [Liu and Lieberman04, Lee05]
 - numerology how to guide
 - Goal Oriented Search Engine (GOOSE)
 - yet, the acquisition of goals has proven to be difficult

Research Overview

Research Question:

- *If and How* search query logs can be utilized to overcome the problem of acquiring knowledge about human goals?

Following an exploratory research style, we intend to show:

- contain a small but interesting number of user goals
- Separation by automatic methods
- Difference of search query goals with regard to goals in other corpora (43Things.com)

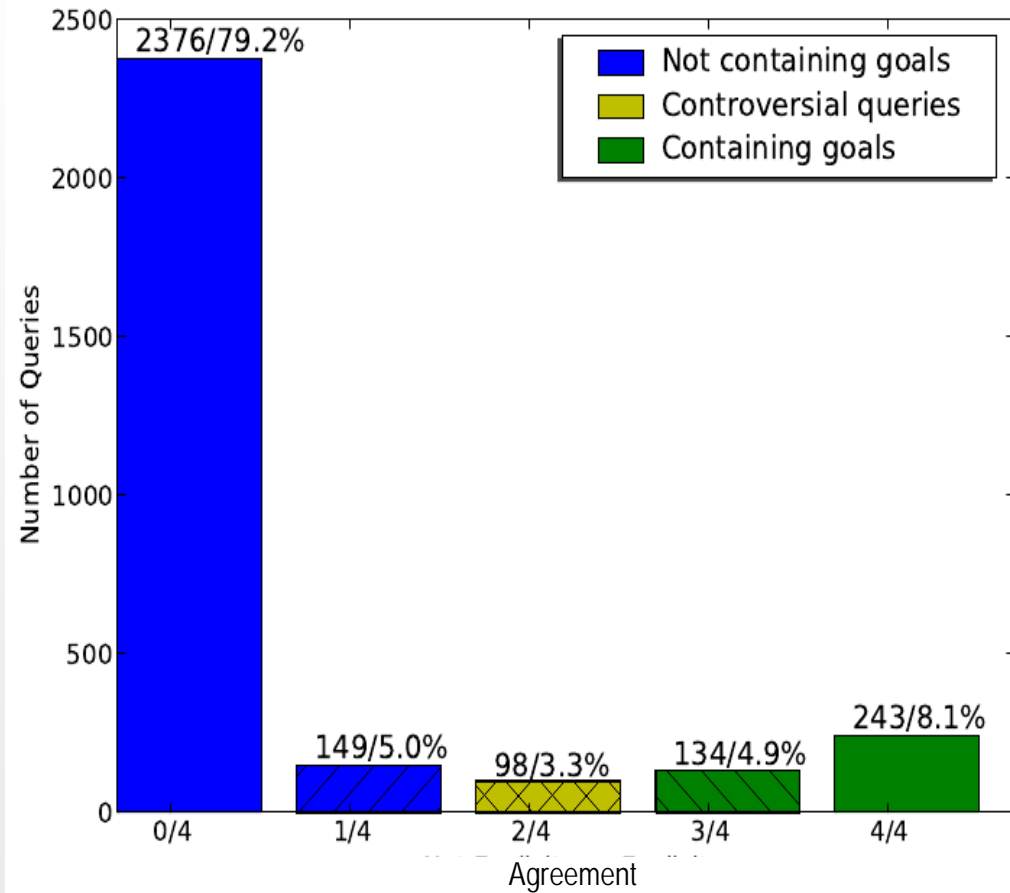
Expected Results:

- Knowledge about the automatic acquisition of goals out of search query logs
- Knowledge about the nature of goals extracted from search query logs

Results of Human Subject Study(1)

- 4 independent raters
- labeled 3000 queries

<i>Pair</i>	κ
A - B	0.86
A - C	0.87
A - D	0.88
B - C	0.83
B - D	0.84
C - D	0.87



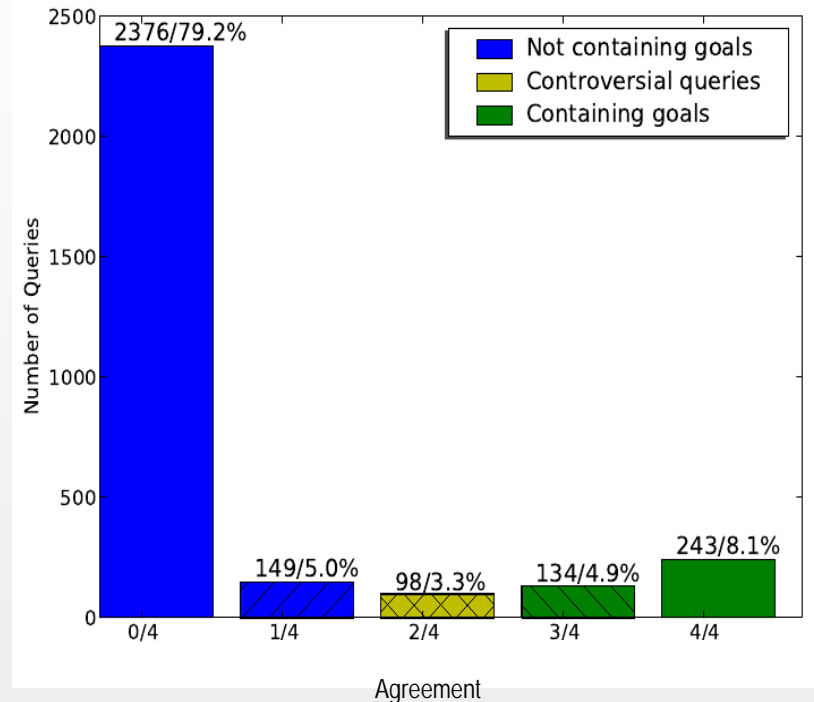
Queries Not Containing Goals → Queries Containing Goals

Results of Human Subject Study(2)

Examples:

- bug killing devices
- mothers working from home
- how to lose weight

- Classes appear to be separable
- → Motivates an automatic approach



Queries Not Containing Goals → Queries Containing Goals

Experimental Setup

- AOL search query log [Pass06]
- ~ 20 million search queries
- recorded between March 1 and May 31 (2006)
- ethical issues

- pre-processing steps to reduce noise
- → pre-processed set comprises 5.405.547 queries

- labeled queries from the human subject study were utilized as training examples (controversial queries were omitted)

Classification Approach(1)

- Feature Types
 - Set of Words
 - stop word removal
 - Part-Of-Speech Trigrams
 - Maximum Entropy Tagger
 - trained on the Wall Street Journal Corpus

Example: Query: “buy a car” → buy/VB a/DT car/NN

Set of words: { buy, car }

Part-of-Speech Trigrams: \$ VB DT NN \$ → { \$ VB DT, VB DT NN, DT NN \$ }

Classification Approach(2)

- Linear Support Vector Machine [Dumais98]
 - Robust and effective in the area of text classification
 - Weka Machine Learning Toolkit [Witten05]
 - No feature selection

- Performance:
 - 10 trails – 3 fold Cross Validation
 - Values averaged
 - Precision, Recall and F1-Measure for the class: “queries containing goals”

Precision	Recall	F1 – Measure
0.77	0.63	0.69

Result Set

- Applying the learnt classifier results in:
 - Result set containing 118.420 entries
 - 97.454 (82,3%) of them are unique

Nr.	Goal Instance	#Users	Nr.	Goal Instance	#Users
1	add screen name	194	11	cancel aol account	46
2	create screen name	98	12	check my computer	41
3	rent to own	85	13	skating with celebrities	40
4	listen to music	78	14	discover credit card	37
5	pimp my ride	64	15	pimp my myspace	34
6	pimp my space	61	16	change my password	33
7	assist to sell	57	17	how to gain weight	32
8	wedding cake toppers	53	18	enterprize car rental	31
9	cancel aol service	50	19	manage my account	30
10	“ <u>deleted</u> ”	47	20	trick my truck	30

Comparative Evaluation

- Comparison with 43Things.com (Goal Corpus)
 - Social Networking site, where users can share lists of goals they want to achieve
 - sample of 36.000 entries

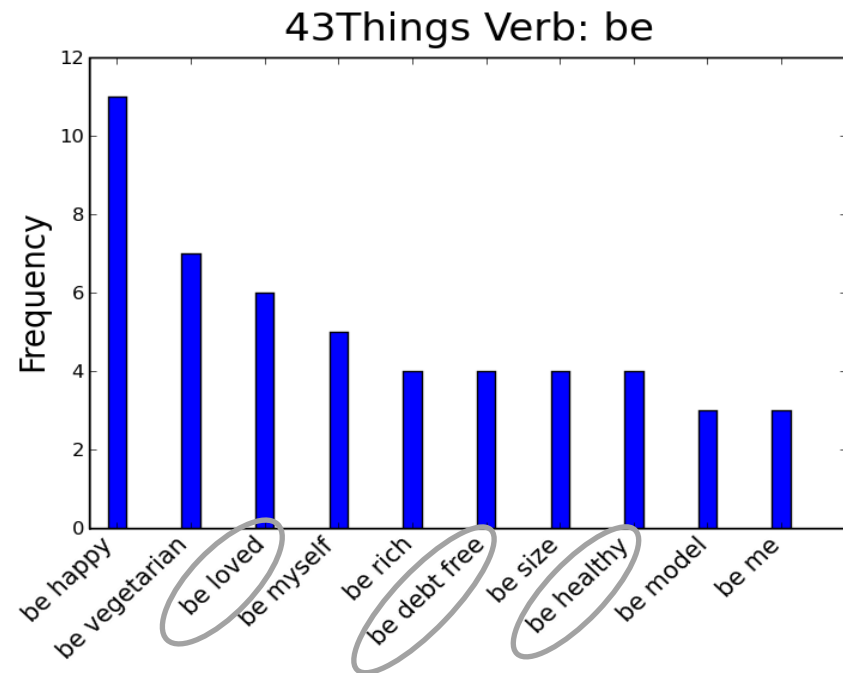
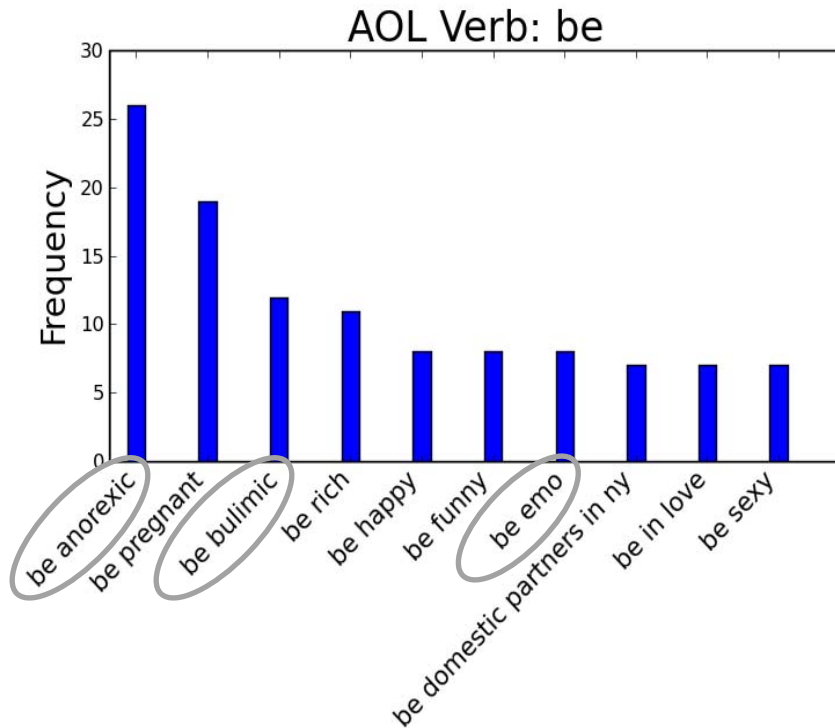
- We are interested, if and how the two datasets differ:
 - Nature of goals
 - Scope of goals

- Perform Qualitative Analysis
 - by examining most frequent entries in both data sets (verbs, nouns, goal instances)

Verbs in AOL vs. 43Things

- Top N most frequent verbs of both goal corpora
- Observations:
 - AOL verbs seem to deal with technical issues
 - 43Things contains more verbs reflecting social activity

#Verbs	AOL	Overlap	43Things
10	buy, listen, sell, use, play	make, find, get, do, learn	be, go have, read, see
50	listen,change, look, move,add, remove, <u>clean</u> , <u>install</u> , <u>apply</u> , draw, put, are, set, convert, rent, tell, <u>fix</u> , <u>pimp</u> , wed, check, cook, <i>deleted</i>	get, be, learn, go,make, have,do, read, see, find, buy, take,write, start,stop, eat, want, keep, create, build, play,use, lose, is, grow, know, sell	s, become, meet, finish, live, watch, run, give, <u>spend</u> , try, own, <u>improve</u> , love, <u>organize</u> , <u>save</u> , <u>speak</u> , <u>join</u> , <u>visit</u> , attend, ride, let, work, am



■ Observations:

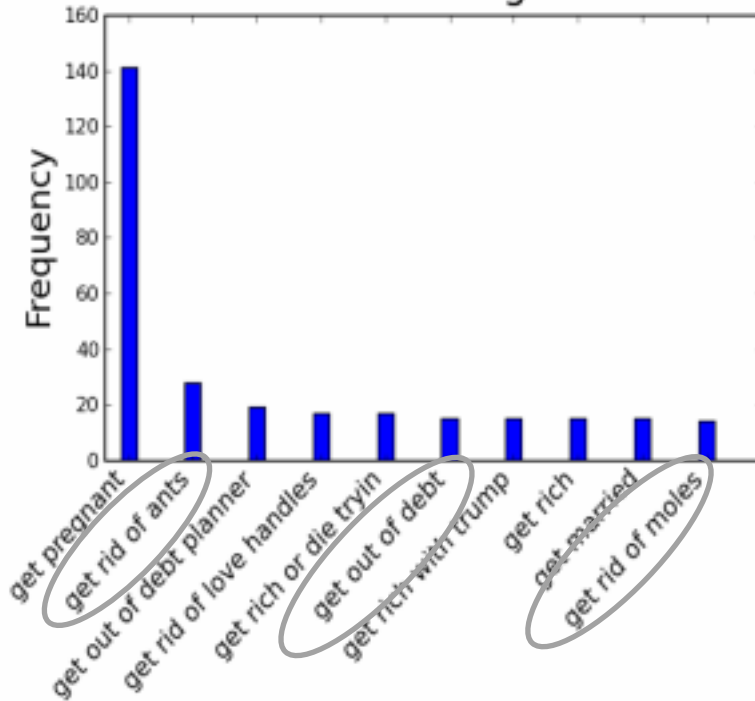
- AOL goals seem to deal with health related issues :

be anorexic, be bulimic, be emo

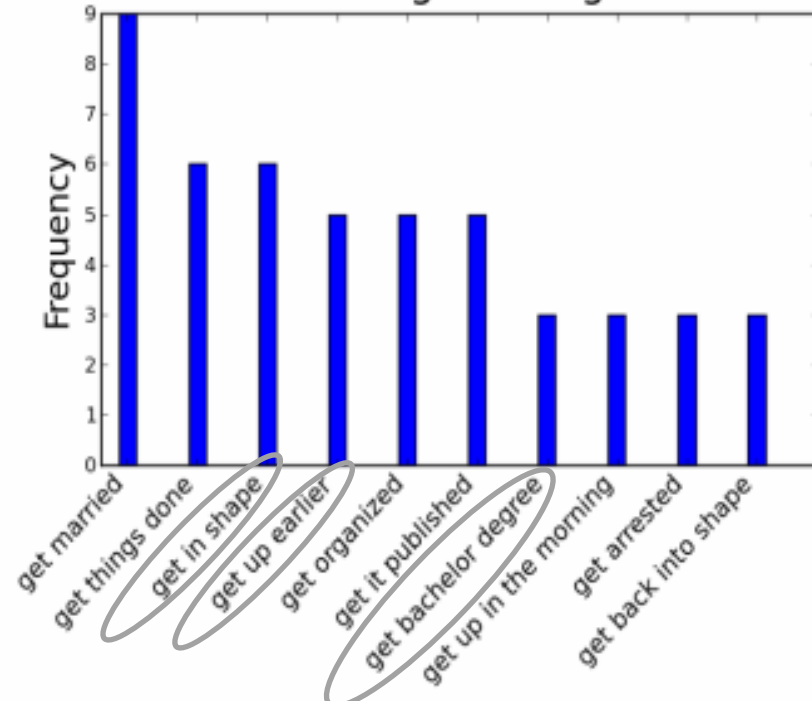
- 43Things goals appear to exhibit a more positive sentiment:

be loved, be debt free, be healthy

AOL Verb: get



43Things Verb: get



■ Observations:

- Users seem to have different time frames in mind - AOL Users often appear to seek immediate answers:

get rid of ants, get out of debt, get rid of moles

- 43Things Users do not seem to underlie these time constraints:

get in shape, get up earlier, get bachelor degree

Contribution

- automatically extracting user goals from search queries seems feasible to a certain extent
- examination of query instances rather than a high level categorization [Broder02]
- search query logs appear to be a promising resource to acquire human goals in an automatic way as opposed [Lieberman07] where human engagement is required

Discussion

- Does the knowledge about a user's search intent allow an improvement concerning the retrieval task?
 - only a very small percentage of queries contain user goals
 - [He07] already attempt to predict user goals from search queries

- Is it likely that users are going to change their attitude about expressing their latent search intent in an explicit way.

Thanks for your attention!

Questions and Discussion



References(1)

- [Dumais98] S.Dumais, J.Platt, D.Heckerman, and M.Sahami. "Inductive learning algorithms and representations for text categorization". In: Proceedings International Conference on Information and Knowledge Management, New York, NY, USA, ACM Press, pp 148-155, 1998
- [He07] K.Y. He and Y.S. Chang and W.H. Lu. Improving Identification of Latent User Goals through Search-Result Snippet Classification. WI '07: Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence, 683-686, IEEE Computer Society,2007.
- [Lee05] U. Lee and Z. Liu and J. Cho. Automatic identification of user goals in Web search. WWW '05: Proceedings of the 14th International World Wide Web Conference, 391--400, ACM Press,New York, NY, USA,2005.
- [Lieberman07] H. Lieberman and D.A. Smith and A. Teeters. Common Consensus: a web-based game for collecting commonsense goals. In Proceedings of the Workshop on Common Sense and Intelligent User Interfaces held in conjunction with the 2007 International Conference on Intelligent User Interfaces (IUI 2007), 2007.
- [Liu and Lieberman02] H. Liu and H. Lieberman and T. Selker. GOOSE: A Goal-Oriented Search Engine with Commonsense. AH '02: Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, 253--263, Springer-Verlag,London, UK,2002.
- [Pass06] G. Pass and A. Chowdhury and C. Torgeson. A picture of search. Proceedings of the 1st International Conference on Scalable Information Systems, ACM Press New York, NY, USA,2006.

References(2)

- [Rose04] D. E. Rose and D. Levinson, Understanding user goals in web search. In Proc. of WWW 2004, May 17-22, 2004, New York, USA, 2004
- [Witten05] I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed., ser. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, June 2005.