Knowledge Management on the Web: Do Search Query Logs Capture Knowledge about Common Human Goals?

Markus Strohmaier, joint work with students Mark Kröll, Christian Körner

Knowledge Management Institute
Graz University of Technology, Austria

e-mail: markus.strohmaier@tugraz.at
web: http://www.kmi.tugraz.at/staff/markus

Markus Strohmaier 2010
Common Human Goals
Traces of User Intent on the Web

A) Search Query

B) Question Answering

C) Human Life

D) Short Messages
“The aggregate results of every search ever entered, every result list ever tendered, and every path taken as a result. […]

This information represents […] a place holder for the intentions of humankind - a massive database of desires, needs, wants, and likes that can be discovered, subpoenaed, archived, tracked, and exploited to all sorts of ends.

**Search as Data**
- Primary means of expressing intent on the web
- Americans conducted **11.8 billion searches** at the core search engines in **July 2008** [Comscore 2008]
  - Very large datasets, up-to-date information, strong user engagement

**Search and User Intent**
- 20% **Navigational**, 48% **Informational** and 30% **Transactional** search intent [Broder 2002, also Rose and Levinson 2004]

**“When users search the web, they think”**
- Users formulate and reformulate plans when searching
- We can/should tap into the outcome of these cognitive processes and exploit them

Why should we collect knowledge about common human goals?
Goal knowledge is required for:
• goal recognition from context / user actions,
• reasoning about user goals
• generation of action sequences that implement goals (planning).

How can we collect knowledge about common human goals?
• Volunteer-based: Openmind/ConceptNet [Singh 2002]
• Text-based: [Tatu 2005]
• Game-based: CommonConsensus [Lieberman 2007]

Common Human Goals
• Describe a common, plausible state of affairs that a human agent may want to achieve or avoid.
Research Overview

**A Characterization Study:*** How useful is the so-called “Database of Intentions” for capturing common human goals?

Following an *exploratory* research style, we aimed to answer:

1: Do Query Logs contain knowledge about common human goals?
2: If they do, what is the nature of common human goals shared by ConceptNet and two large Query Logs?
3: How do goals contained in ConceptNet and Query Logs differ wrt. scope?
4,5: Can goals contained in Query Logs be used to *refine* and *expand* ConceptNet?
Vision: Intelligent Agents Acting Upon User Intent

ConceptNet 2.1:

Challenges:
1. Goal Acquisition Problem
2. Goal Coverage Problem

Out of >2 mio assertions, ~50,000 – 70,000 refer to common human goals
## Comparative Study Setup

### Goals from ConceptNet
- ConceptNet 2.1
- more than 20 relation types
- POS tagging

#### Acquisition:
- satisfy additional constraints:
  - “MotivatedByGoal”, “UsedFor”, “CapableOf”
  - \( \text{VB}+ \text{NN}+ \)

#### Result:
- \(|C| = \sim 68,000\) goals

### Goals from Search Query Logs
- two large search logs from 2006 (AOL and Microsoft Research)
- Total of \(\sim 35\text{mio}\) queries
- \{UserID, Query, Timestamp, (ItemRank, URL)\}  

#### Acquisition:
- Using an existing goal acquisition method [26]

#### Result:
- \(|S| = \sim 115,000\) goals
  - (77% precision)
Different Degrees of Explicitness in Search Queries

M. Strohmaier, P. Prettenhofer, M. Lux, Different Degrees of Explicitness in Intentional Artifacts - Studying User Goals in a Large Search Query Log, CSKGOI'08 International Workshop on Commonsense Knowledge and Goal Oriented Interfaces, in conjunction with IUI'08, Canary Islands, Spain, 2008.

Example Queries:

- car Miami
- car Miami dealer
- buy a car in Miami
- buy a used car in Miami
- get loan to buy a used car in Miami

**Definition:** A search query is regarded to contain an explicit user goal whenever the query
1) contains at least one verb and
2) describes a plausible state of affairs that the user may want to achieve or avoid in a recognizable way.

<table>
<thead>
<tr>
<th>Queries containing explicit statements of goals</th>
<th>Queries not containing explicit statements of goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;sell my car&quot;</td>
<td>&quot;Mazda dealership&quot;</td>
</tr>
<tr>
<td>&quot;play online poker&quot;</td>
<td>&quot;online games&quot;</td>
</tr>
<tr>
<td>&quot;find home to rent in Florida&quot;</td>
<td>&quot;Miami beach houses&quot;</td>
</tr>
<tr>
<td>&quot;passing a drug test&quot;</td>
<td>&quot;drug test&quot;</td>
</tr>
<tr>
<td>&quot;raising your credit score&quot;</td>
<td>&quot;credit cards&quot;</td>
</tr>
</tbody>
</table>

*Table: Example Queries*
Do Query Logs contain Goal Formulations?  
A Method for Goal Acquisition


Step 1: Part-of-Speech tagging queries
• Only queries with query length >2 (removes ~ 60% of queries)
• Feature Types
  – Set of Words (stop words removed)
  – Part-Of-Speech Trigrams
    • Maximum Entropy Tagger
    • trained on the Wall Street Journal Corpus

Step 2: Supervised Learning of Syntactical Goal Features
• Part-of-speech (POS) trigrams

"buying/VBG a/DT car/NN"

$ $ VBG DT NN $ $ $ $ VBG, $ VBG DT, VBG DT NN, DT NN $, NN $$

• Support Vector Machine trained on the feature vectors using 10-fold cross-validation
• Result: 77 % precision, 63% recall
Results & Most Discriminative Features

From previous work: [Strohmaier 2007, 2008]

• 1-3% of queries contain actual expressions of goals

• Search can be understood as a traversal of goal graphs

<table>
<thead>
<tr>
<th>Nr.</th>
<th>POS</th>
<th>SOW</th>
<th>#</th>
<th>Feature</th>
<th>Example Matching Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>126</td>
<td>$ WBE TO</td>
<td>($ where to) find roses in Georgia</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>130</td>
<td>WBE TO VB</td>
<td>[how to fix] your cold</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>83</td>
<td>TO VE NN</td>
<td>drink milk [to lose weight]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>41</td>
<td></td>
<td>buy now pay later jewelry</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>58</td>
<td>VB NN NN</td>
<td>[find property values] calculator</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>20</td>
<td></td>
<td>find [an old friend] for free</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>36</td>
<td>TO VE JJ</td>
<td>I want [to download picture] message</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>X</td>
<td>27</td>
<td></td>
<td>make your own parable</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>X</td>
<td>52</td>
<td>$ VB NN</td>
<td>[$ find browser] in Georgia to start it</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>X</td>
<td>29</td>
<td>VB NN IN</td>
<td>[borrow money from] Donald Trump</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>X</td>
<td>12</td>
<td>TO VE PRP</td>
<td>how [to copyright your] photos</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>X</td>
<td>14</td>
<td>WRB VP PRP</td>
<td>my hair turned orange [how do I] fix it</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>X</td>
<td>26</td>
<td>TO VE NN</td>
<td>what [to eat] Mexican</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>X</td>
<td>28</td>
<td>VB NN NN</td>
<td>[make business cards]</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>X</td>
<td>19</td>
<td>TO VE DT</td>
<td>teach yourself [to play the] piano</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>X</td>
<td>9</td>
<td>VB PRP JJ</td>
<td>how to [get yourself sick]</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>X</td>
<td>43</td>
<td>TO VE IN</td>
<td>places [to stop in] Gothenburg</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>X</td>
<td>8</td>
<td></td>
<td>install</td>
<td>install Microsoft windows 2000</td>
</tr>
<tr>
<td>19</td>
<td>X</td>
<td>14</td>
<td>VB PRP</td>
<td>[$ save your] old buddy icon</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>X</td>
<td>22</td>
<td>VB PRP NN</td>
<td>how to [obtain a passport]</td>
<td></td>
</tr>
</tbody>
</table>
Search Query Log Analysis - Results

M. Strohmaier, P. Prettenhofer, M. Kroell, Acquiring Explicit User Goals from Search Query Logs
In Proceedings of the International Conference on Web Intelligence WI’08, Agent and Datamining Interaction Workshop, Sydney, Australia, 2008

Commonsense goals in the „tail“

Table 8: The 10 most frequent verb phrases containing the verbs “get”, “make”, “change”, and “be” are listed.

Goals marked with (*) are also included in ConceptNet Commonsense Knowledge Base v2.1
[H. Liu and P. Singh 2004]
Existence of Common Human Goals

Do Search Query Logs contain knowledge about common human goals?

|C| ~68,000 goals  |S| ~115,000 goals

Simple Goal Matching Algorithm:

- Stopword removal, Stemming, Bag-of-Words

<table>
<thead>
<tr>
<th>ConceptNet Goals</th>
<th>Search Query Log Goals</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>make paper airplanes</td>
<td>how to make paper airplanes</td>
<td>yes</td>
</tr>
<tr>
<td>get in shape</td>
<td>getting into shape</td>
<td>yes</td>
</tr>
<tr>
<td>we buy houses</td>
<td>buy a house</td>
<td>yes</td>
</tr>
<tr>
<td>we buy houses</td>
<td>purchase a house</td>
<td>no</td>
</tr>
<tr>
<td>make money</td>
<td>make more money</td>
<td>yes</td>
</tr>
<tr>
<td>make money</td>
<td>make online money</td>
<td>No</td>
</tr>
</tbody>
</table>

Result:

- **Little Overlap!** ~2300 ConceptNet goals and ~3100 search query log goals (occurrences) that produced positive matches.

Does this mean that there is little or much to gain from acquiring goals from search query logs?
Levin Verb Classes

A classification of over 3,000 English verbs according to shared meaning and behavior into 193 coherent verb classes.

Example: Verb Class „Admire“ (31.2):
Abhor, admire, adore, appreciate, cherish, enjoy, fancy [excerpt]

We selected 15 verb classes and its alterations for further analysis:

Admire, Build, Cable, Create, Eat, Feel, Give, Go, Learn, Obtain, Perform, Prepare, Search, Tell, Turn

How are goals distributed across these classes?
Commonalities Between Datasets

What is the nature of goals shared by ConceptNet and Query Logs?

Levin Verb Classes [Levin 1993]

- "Build" (cook, develop, hack, make)
- "Obtain" (buy, get, order, win)
- "Perform" (play, produce, sing, write)

Figure 1: Distribution of goals over Selected Levin's Verb Classes [Levin 1993] in the Intersection Set.

Result:
- “Build”, “Obtain” and “Perform” are types of goals that dominate the intersection set.
- Evidence of (selected) types of common human goals in search query logs

Markus Strohmaier 2010
Differences Between Datasets

Goals contained in ConceptNet and Search Query Logs differ w.r.t. scope, and if so - how?

Selected Levin’s word classes:

- **“Build”** cook, develop, hack, make
- **“Obtain”** buy, get, order, win
- **“Admire”** love, hate, enjoy, regret
- **“Want”** crave, hope, long, wish, yearn
- **“Feel”** sense, smell, taste, scent
- **“Go”** climb, crawl, run, descend

Result:
- ConceptNet exhibits a more uniform distribution of verb classes (broad focus)
- Distribution of goal types from Search Query Logs are more skewed (narrow focus)

Markus Strohmaier 2010
Refining ConceptNet using Query Logs
Illustration by Example

Simple refinement idea:
1) stem all words
2) find search goals that contain all stems from ConceptNet goals
3) require search goals to have additional stems (\ADJ*)

Avg. of 3 refinements per ConceptNet goal!

<table>
<thead>
<tr>
<th>ConceptNet Goal</th>
<th>List of Goal Refinements From Query Logs</th>
</tr>
</thead>
<tbody>
<tr>
<td>now buy this car</td>
<td>buy new car (20), buy a cheap car (2), buying rental cars (2), buy electric car (2), buy a used car (19), buy old cars (3), buying wise car (1)</td>
</tr>
<tr>
<td>finding friends</td>
<td>to find old friends (4), find high school friends(1), find lost friend (11), find best friends (1), find elementary school friend (1), find free online friends (1), find past military friends (1)</td>
</tr>
<tr>
<td>writing a paper</td>
<td>write an argumentative paper (1), write an informative paper (1), write an autobiographical paper (1), write a narrative paper (1)</td>
</tr>
<tr>
<td>cutting your hair</td>
<td>cutting my own hair (1), cut short hair (3), cut black hair (1), cutting long hair (1)</td>
</tr>
<tr>
<td>feeding the baby</td>
<td>feeding a newborn baby (1)</td>
</tr>
<tr>
<td>train a dog</td>
<td>train an abused dog (1), train a deaf dog (1)</td>
</tr>
</tbody>
</table>

Figure 1: Exemplary Refinement Candidates for Selected ConceptNet Goals.

Result:
- By example, we have shown that search query logs can add a significant number of relevant refinements to commonsense knowledge bases such as ConceptNet.
Expanding ConceptNet using Query Logs
Illustration by Example

“Do you think that a person’s motivation could be M when performing action A?”

<table>
<thead>
<tr>
<th>Goal from Search Query Logs</th>
<th>List of ConceptNet Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>make some money quickly</td>
<td>wait tables, go to work, work the box office, serve customers, tell a story, get a contract, buy a house, apply for a job, pass a course</td>
</tr>
<tr>
<td>make new friends in your area</td>
<td>meet interesting people, meet people</td>
</tr>
<tr>
<td>find credit information</td>
<td>surf the net, surf the web, use a computer</td>
</tr>
<tr>
<td>ways to gain weight</td>
<td>eat ice cream</td>
</tr>
<tr>
<td>lose maximum weight fast</td>
<td>go jogging, eat healthily, release your energy, go for a run, play sports, get exercise, get some physical activity, eat vegetables</td>
</tr>
</tbody>
</table>

Result:
- 77 out of 120 goals were regarded reasonable motivations (precision of 64%)
- Cohen’s kappa $\kappa$ 0.36 (fair agreement)
Goal Dataset is Available for Research

SearchGoalNet:
• Acquired from the AOL 2006 dataset
  • Contains > 70,000 inter-linked goals (untyped, weighted relations)
  • precision 77%
• Data accessible via a RESTful API
  • /Predict/
  • /RelatedGoals/
  • /URLofGoal/
• Demo (requires key)
  • http://level6.know-center.tugraz.at:8080/GoalNetService/

Contact: markus.strohmaier@tugraz.at
The Problem

Example: A query that probably many prospective college student in the US performs once in his life.

Optimized for navigational queries

Estimating User Intent
Goal Estimation Task

**Input:** 1 to m queries

Example:
- English cottage

Output: weighted list of n goals

Example:
- find house plans (0,400)
- design my own house (0,325)
- design your own house plans (0,300)
- build your own house (0,300)
- build your own english cottage (0,300)

Queries not containing goals

Queries containing goals

**Understanding user intent:**
A major barrier in search
**Intentional Query Suggestion:** Relating queries not containing explicit user goals with queries that contain explicit user goals

![Diagram showing query log neighbourhood of explicit user goals](image)

Excerpt of the AOL search query log. User Ids are omitted. Queries are sorted in order of occurrence.

\[
\vec{g} = f(d, i, g(w))
\]
The Graph Construction Process / 1

- **Idea**: use tags to build a bipartite graph
  - First mode: goals
  - Second mode: tags
The Graph Construction Process / 2

- We fold the bipartite graph into a unipartite graph consisting only of goals

Folding: multiplying the affiliation matrix $A$ with its transpose ($=A^*A^T$)
Measures of Goal Similarity

- **Neighbourhood**-based similarity \( N(g_1, g_2) \)

<table>
<thead>
<tr>
<th></th>
<th>Goal ( g_1 )</th>
<th>Goal ( g_2 )</th>
<th>Query ( q_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>cottage</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>world</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>plans</td>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>old</td>
<td>3</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>build</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>house</td>
<td>6</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Text**-based similarity \( T(g_1, g_2) \)
  
  „Buy a car“ <-> „car“

- **Combined** similarity
  
  \( CS(g_1, g_2) = N(g_1, g_2) \times \alpha + T(g_1, g_2) \times (1-\alpha) \)

**Similarity coefficients:**

- Dice‘s coefficient \( D(X, Y) \)
  
  \[
  D(X, Y) = \frac{2 |X \cap Y|}{|X| + |Y|}
  \]

- Jaccard‘s coefficient \( J(X, Y) \)
  
  \[
  J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}
  \]

- Cosine coefficient \( C(X, Y) \)
  
  \[
  C(X, Y) = \frac{|X \cap Y|}{\sqrt{|X| \times |Y|}}
  \]

- Overlap coefficient \( O(X, Y) \)
  
  \[
  O(X, Y) = \frac{|X \cap Y|}{\min(|X|, |Y|)}
  \]

**Counting measure \(| . |\) gives the size of the set.**
Goal Estimation Task

Input: 1 to m queries

Example:
• English cottage

Underlying goal?

Output: weighted list of n goals

Example:
• find house plans (0,400)
• design my own house (0,325)
• design your own house plans (0,300)
• build your own house (0,300)
• build your own english cottage (0,300)

Goal hypotheses

<table>
<thead>
<tr>
<th>Goal g1</th>
<th>Goal g2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>
Demo
Goal Estimation

Enter your short query

Examples: "car", "college" or "house"

Predict my goal →

Graph Similarity: OVERLAP
Text Similarity: OVERLAP
Alpha: 0.5
Ranking = graph * alpha + text * (1-alpha)

Select Network

Results

Your query: house
Alpha is set to: 0.5
The chosen graph similarity is: ADAPT_OVERLAP
The chosen text similarity is: ADAPT_OVERLAP
Found following goals and probabilities:

- save my house (0.667) G: 0.500 T: 0.167
- painting the house (0.667) G: 0.500 T: 0.167
- insure my house (0.667) G: 0.500 T: 0.167
- see my house (0.667) G: 0.500 T: 0.167
- sell your house (0.667) G: 0.500 T: 0.167
- buy house windows online (0.625) G: 0.500 T: 0.125
- building your own house (0.625) G: 0.500 T: 0.125
- how to paint your house (0.600) G: 0.500 T: 0.100
Excerpt of Goal Graphs Constructed from Search Query Logs

Based on the AOL Search Query Log, ~20mio queries [Pass 2006]
Goal Graph with ~70,000 goals, precision of entries: ~70%

Available via a RESTful Web-Service
SearchGoalNet – A RESTful API
joint work with C. Körner

Example: .../GoalNetService/goals/car?id=dummy
Replace dummy with your username

/goals?id=dummy returns all goals in json format (Value: "Goals", Key: Array of Strings representing the goals)
/tags?id=dummy returns all tags in json format (Value: "Tags", Key: Array of Strings representing the tags)

/goals/making quick cash?id=dummy returns all tags connected to the goal in json format (Key: tag, Value: Weight, Key: tag, Value: Weight, etc.)
/tags/car?id=dummy returns all goals connected to the tag in json format (Key: goal, Value: Weight, Key: goal, Value: Weight, etc.)

/rawRelatedGoals/making quick cash?id=dummy returns all goals which share one or more tags with the goal in json format (Key: goal, Value: array of key, value pairs - goal and number of shared tags)
/predict/car dealer?id=dummy&text=DICE&graph=DICE&alpha=0.5 returns all goals which are predicted by the goal prediction in json format (Key: query, Value: array of key, value pairs - goal and calculated weight)
/relatedGoals/making quick cash?id=dummy&text=DICE&graph=DICE&alpha=0.5 returns all goals which are predicted by the goal prediction in json format (Key: query, Value: array of key, value pairs - goal and calculated weight)

/URLofGoal/how to make money from stocks?id=dummy returns all URLs which are associated to a given goal in json format (Key: goal, Value: all urls in apostrophes, separated by commas)

Available Similarities:
- DICE
- COSINE
- OVERLAP
- ADAPT_OVERLAP
- JACCARD

The provided alpha range has to be between 0.0 and 1.0.
Prototype: Intentional Query Expansion

Greasemonkey: joint work with F. Wörister

Evaluation Setup:
5 human annotators
35 short queries each (1-2 words)

Evaluation Task:
“Would goal G be a goal of a user who issues query Q?”

Answer 1: Yes, G is a plausible or potential goal for Q
Answer 2: No, G is not a relevant goal for Q

Evaluation Results:
Pairwise average interrater agreement: $\kappa = 0.67$
Precision: 72% (based on a majority vote among 5 annotators)
Evaluation Design

Model:
- AOL Search Query Log
- 113,420 d-neighbourhood search sessions that contain an explicit user goal

Model Parameters
- Neighbourhood d, Weight function f(w), token selection i

HoldOut:
- 500 random search sessions
- Containing an explicit goal that occurs at least twice in the dataset

Estimation:
- Input: HoldOut Sessions \ Explicit User Goals
- Estimation Parameters
  - Neighbourhood d
  - Weight function f(w)
  - Window (Pre|Post|Pre&Post Queries)
  - Similarity Coefficient (J|D|O|C)
- Output: Ranked list of estimated goals (estimation list)
- Metrics: Recall, rank of estimated goal (exact string matching)
In how many sessions did our algorithm include the correct goal in the estimation list?

**Input** $d=1$ (Input consists of the previous and/or the succeeding query only)

Text similarity alone yields interesting results, CS improves recall slightly. Input $>1$ does not yield significantly better recall (*not shown*).
500 Evaluated Random Sessions: Rank of Correct Goal in Prediction Sets

At what rank did our algorithm position the correct goal in the estimation list?

**Similarity:**

**Text**

**Combined**

**Graph**

Observation: Dependant on the parameters, the mode lies between rank 1 and 5.

The algorithm estimated the correct goal on rank 1 in ~30 out of 500 cases.

\[ CS(g_1, g_2) = N(g_1, g_2) \times \alpha + T(g_1, g_2) \times (1-\alpha) \]
500 Evaluated Random Sessions: Median Rank of Correct Goal in estimation Sets

What is the **median rank** of correct goals in the estimation list?

By just looking at a single query (PRE or POST), and using TEXT similarity (with d=3, i=1), the algorithm predicts the correct goal:

- on rank 5 (median) with a recall of 35%
- on rank 9 (median) with a recall of 45%

---

Markus Strohmaier 2010
Conclusions

Contributions:
• A characterization study of the usefulness of search query logs for acquiring knowledge about common user goals

• By example, we showed how, and in which areas, query logs could add relevant knowledge to commonsense knowledge bases

• Query logs as narrow, but abundant and (to some extent) useful source of knowledge about common user goals

• limitations of utilizing query logs

Issues and Further Work:
• Availability and Scope of Search Query Logs
• Inherent difficulty of evaluating commonsense knowledge (defeasibility)
• Robust methods to relate goals from Search Query Logs and ConceptNet
QUESTIONS?

See you next week!
Papers

M. Strohmaier, P. Prettenhofer, M. Lux, Different Degrees of Explicitness in Intentional Artifacts - Studying User Goals in a Large Search Query Log, CSKGOI'08 International Workshop on Commonsense Knowledge and Goal Oriented Interfaces, in conjunction with IUI'08, Canary Islands, Spain, 2008. (pdf)


M. Strohmaier, M. Kroell, Studying Databases of Intentions: Do Search Query Logs Capture Knowledge about Common Human Goals?, The Fifth International Conference on Knowledge Capture (K-CAP'09), Sep 1-4, Redondo Beach, California, USA, 2009. (Acceptance rate: 21/81, 25.92% quota) (pdf)

M. Kroell, M. Strohmaier, Analyzing Human Intentions in Natural Language Text, The Fifth International Conference on Knowledge Capture (K-CAP'09), Sep 1-4, Redondo Beach, California, USA, 2009. Best Poster Award out of 21 submissions (4.76% quota) (PDF paper, PDF poster)