

707.000
Web Science and Web Technology
„User Intentions and Intentional Structures
on the Web“

Markus Strohmaier

Univ. Ass. / Assistant Professor
Knowledge Management Institute
Graz University of Technology, Austria

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

Overview

Poll:

http://kmi.tugraz.at/staff/markus/courses/707.000_web-science/

Overview

Agenda

A glimpse at recent and ongoing research at the KMI:

- [How do users express goals on the web? - An exploration of intentional structures in web search.pdf](#)
- A prototype for intentional query expansion
- [Different degrees of explicitness in intentional artifacts - Studying user goals in a large search query log.pdf](#)

Motivation

Search engine vendors are desperate to learn more about users intentions

Peter Norvig, Director of **Google** Research

- understanding users' needs to a greater extent represents an "*outstanding*" research problem.
- Google is currently looking at "*finding ways to get the user more involved, to have them tell us more of what they want.*" [2]

Yahoo's future strategy aims at "driving deeper engagement" with its users through vertical search, personalization, desktop search and social search and believes the increasing sophistication of user search behavior supports the goal (2006)

<http://blogs.zdnet.com/micro-markets/index.php?p=27>

-The Database of Intentions-

John Battelle

- The „Database of Intentions“ is

“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.

Such a beast has never before existed in the history of culture [...].”

Source: <http://battellemedia.com/archives/000063.php>, last accessed on July 3rd, 2007

Selected Research Questions

1. How can we **identify goals** in search query logs?
2. How can we represent search goals in a **semi-formal goal graph**?
3. How can we represent the search process as a **traverse** through the goal graph?
4. How can goals **direct search behaviour**?
5. Is there a difference between **explicit and implicit intentional search queries**, and how can it be identified?
6. How can we construct **large scale goal graphs**?

Example: AOL Search Database

- History
 - Released with the purpose of serving science
 - More than 20 million queries
 - User IDs, timestamps, queries, URLs selected
- Implications
 - “A Face Is Exposed for AOL Searcher No. 4417749”
<http://www.nytimes.com/2006/08/09/technology/09aol.html?ex=1312776000&en=f6f61949c6da4d38ei=5090>
 - „57thiest dumbest moment in business“
 according to Business 2.0 Magazine on CNNMoney:
http://money.cnn.com/magazines/business2/101dumbest/2007/full_list/index.html
- Data online
 - http://sergiorebelo.com/twodotfive/?page_id=25
 - Links to several online services for searching the AOL search database (September 2007)
 - <http://www.aolstalker.com/>
 - „Überwacht die Überwacher“

Motivation

Today,

- many activities on the web are *driven by goals*
(e.g. *buy car, find gift, treat influenza*)
- many services and resources on the web already provide varying support for *achieving goals*
- Users routinely *translate high-level goals* into specific search queries, tag concepts or classification terms

yet

- The web lacks explicit **intentional (i.e. goal-oriented) structures**

Which makes it

- hard for users to explore and express goals
- difficult for technologies to discover, represent, comprehend and reason about users' goals
- difficult for communities to share and discover goals
- ...

Goals and Intentions

Goals are
„desired states of affairs“

Intentions are
„goals + commitment“
[Mylopoulos 2006]

Some high-level examples of potential users' goals on the web:

- Gifts -> create | find | buy | sell | ...
- Wedding -> organize | attend | sabotage | film | ...
- Car -> buy | sell | repair | tune-up | insure | ...
- Influenza -> treat | avoid | eliminate | ...

Structure of Goals

(in Requirements Engineering)

Simplified Assumption:

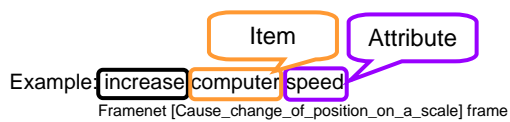
Goal = Verb + Semantic Frame elements
(see, e.g. [Liaskos 2006])

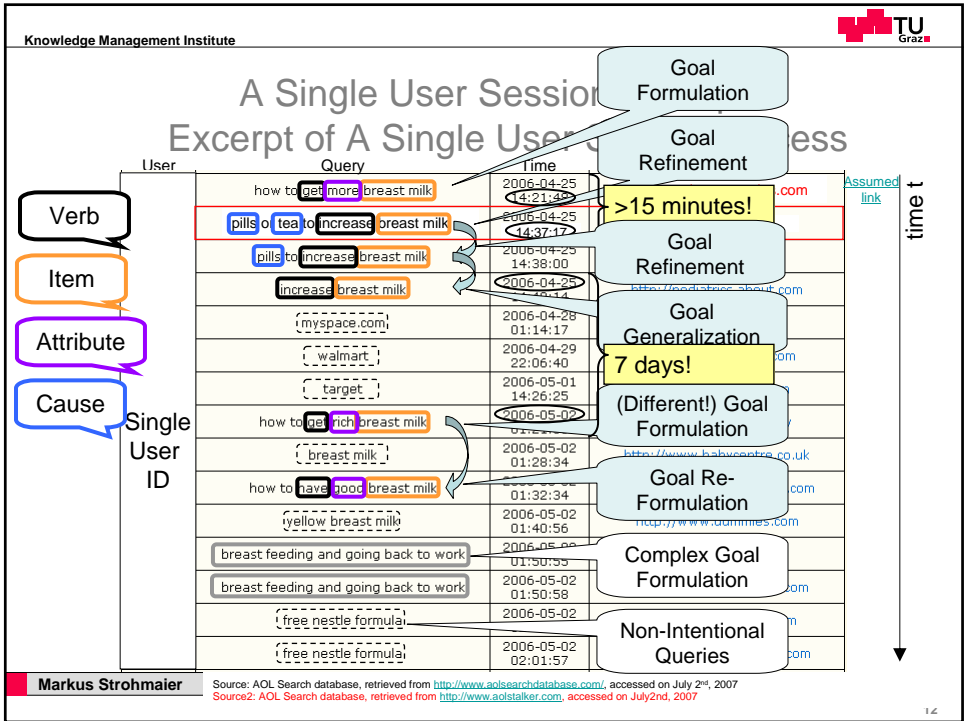
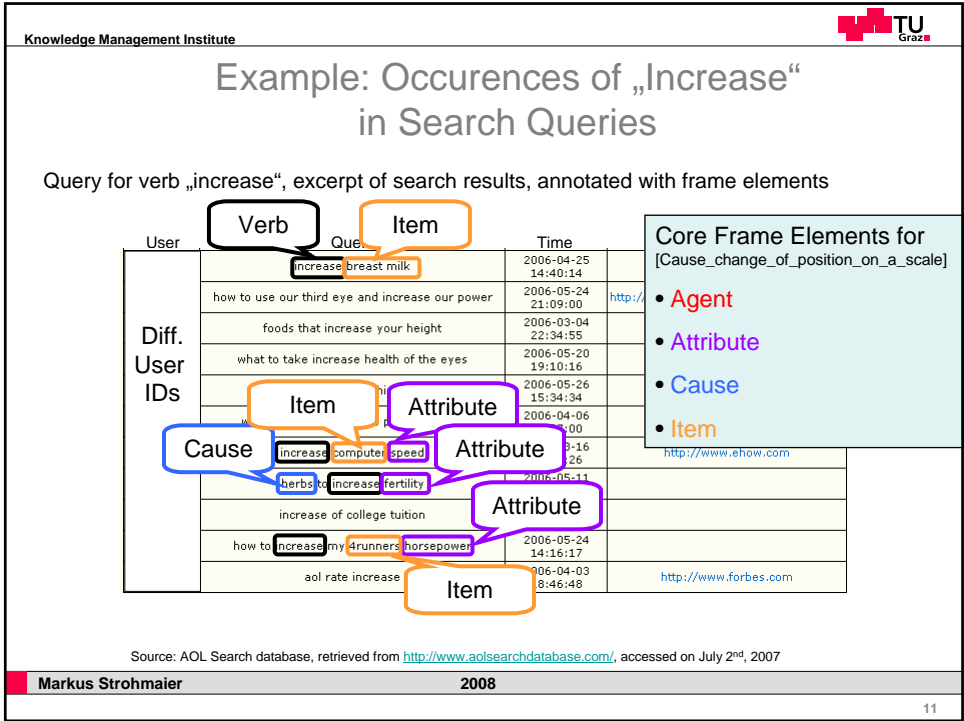
Example verbs: Achieve, Make, Improve, Speedup, Increase, Satisfied, Completed, Allocated, Maintain, Keep, Ensure, Avoid, Know, Monitor, Track, Provide, Supply, ... [Regev 2005]

The Berkeley Framenet Project [Ruppenhofer et al 2006]:

Semantic Frame: “a script-like conceptual structure that describes a particular type of situation, object or event along with its participants and props”

Frame Elements: Each of these elements of a semantic frame is called frame elements





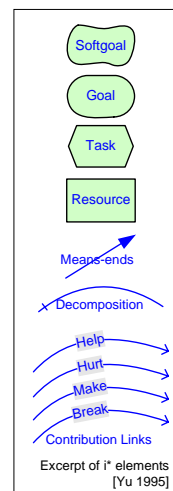
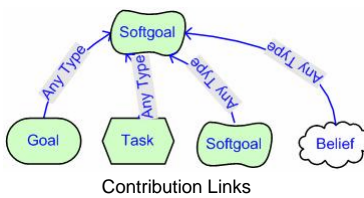
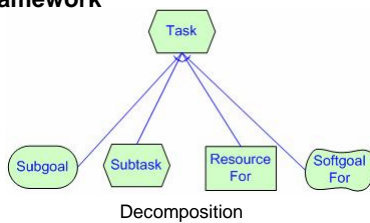
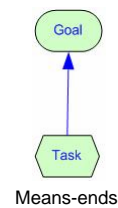
Questions

1. How can we represent the involved goals in a **semi-formal goal graph**?
2. How can we represent the search process as a **traverse** through the goal graph?

Intentional Structures - Goal Models in Requirements Engineering

Excerpts of the i* framework

[Yu 1995]



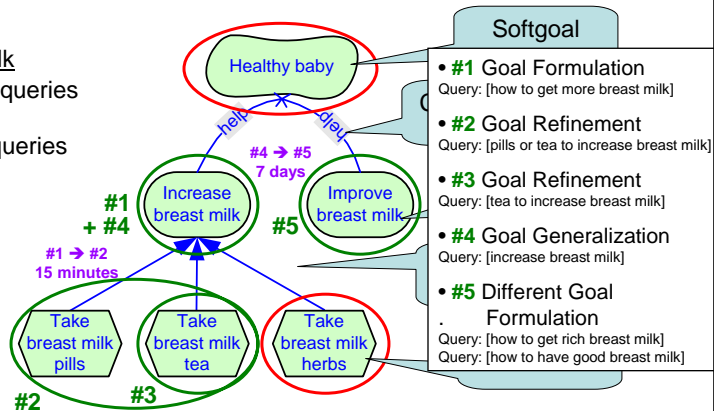
Source: <http://istar.rwth-aachen.de/tiki-index.php?page=iStarQuickGuide>, last accessed on July 3rd, 12007

Questions

1. How can we represent the involved goals in a **semi-formal goal graph**?
2. How can we represent the search process as a **navigation** through the goal graph?

Domain: Breast Milk

- Executed search queries
- Time differences
- Not-executed search queries



Based on the NFR/* Framework
[Chung et al 2000/Yu 1995]

Other web2.0 applications „How to X?“

Examples:

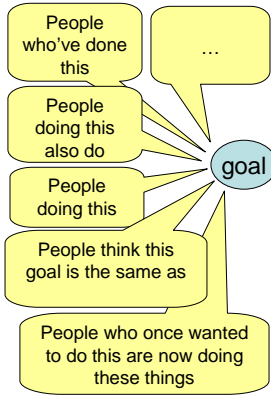
- www.ehow.com / www.wehow.com
how to descriptions & videos
- www.wikihow.com
- www.videojug.com
explanatory videos
- www.viewdo.com
includes “Required Tools” and “Required Parts”
- www.howto.tv
- www.43things.com
- More at http://dir.yahoo.com/Reference/How_To_Guides/



43 Things API

Common XML Objects

- [Person](#)
- [Goal](#)
- [Entry](#)
- [City](#)



Goal Methods:

- [search_goals](#)
- [get_goal_by_id](#)
- [get_goal_by_name](#)
- [get_goals_similarities](#)
- [get_goals_people](#)
- [get_goals_entries](#)

People Methods:

- [search_people](#)
- [search_people_by_email](#)
- [get_person](#)
- [get_persons_completed_things](#)
- [get_persons_entries](#)
- [get_persons_progress_on_goal](#)
- [get_persons_teammates](#)
- [get_persons_neighbors](#)
- [get_persons_tags](#)
- [get_persons_tag_cloud](#)
- [get_subscriptions](#)

Team Methods:

- [get_teams_progress](#)

Entry Methods:

- [get_entry](#)

Tag Methods:

- [search_tags](#)
- [get_tags_goals](#)
- [get_tags_similarities](#)

City Methods:

- [search_cities](#)
- [get_city](#)
- [get_cities_people](#)

Action Methods

- [add_goal_by_id](#) (requires authentication)
- [add_goal_by_name](#) (requires authentication)
- [complete_goal](#) (requires authentication)
- [remove_goal](#) (requires authentication)
- [update_goal](#) (requires authentication)

ATOM Methods:

- [introspection](#)
- [entry](#) (requires authentication)

+ RSS

Source: http://www.43things.com/about/view/web_service_api
last accessed on July 4th, 2007

Selected Research Questions

1. How can we **identify goals** in search query logs?
2. How can we represent search goals in a **semi-formal goal graph**?
3. How can we represent the search process as a **traverse** through the goal graph?
4. How can goals **direct search behaviour**?
5. Is there a difference between **explicit and implicit intentional search queries**, and how can it be identified?
6. How can we construct **large scale goal graphs**?

Result of one of the projects conducted by „Knowledge Technologies“ students
C. Ruggenthaler und A. Haselsberger

In which principal ways can we modify search?

Query modification

Velez et al 1997

- **Scope reduction** (of underspecified queries)
(e.g. refinement, conjunction)
- **Scope expansion** (of overspecified queries)
(e.g. generalization, disjunction)

Result adaptation

- *Adapting the content*
- *Re-ordering the links*
- *Augmenting the links*

e.g. ScentTrails

e.g. GOOSE

What can be adapted in adaptive hypermedia?
<http://www.wis.win.tue.nl/ah94/Brusilovsky.html>

Query Modification

Query Refinement is the incremental process of transforming a query into a new query that more accurately reflects the user's information need. [Velez et al 1997]

- Current implementations e.g. search for "car" on ask.com ->
 - new car, used cars, fast cars, sports cars, car insurance
 - predominately non-intentional

Our idea: **Intentional query expansion** focuses on query expansion exclusively from an intentional perspective:

- E.g. search for "wine crop" ->
 - verbs: protect wine crop, increase wine crop, improve wine crop
- E.g. search for "increase wine crop" ->
 - means: fertilizer to increase wine crop, insecticides to increase wine crop
 - helps: winery be successful, harvest more wine
 - hurts: produce organic wine

→ Support the user in formulating/refining/disambiguating/learning about his/her goals.

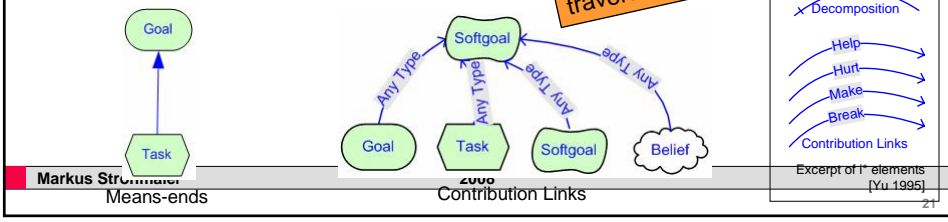
I) What is Intentional Query Expansion?

Definition:

Intentional query expansion is the incremental process of transforming a query into a new query based on **intentional structures found in a given domain**

What are intentional structures?

Intentional query expansion as a guided traversal of goal graphs



Markus Strohmaier
Means-ends

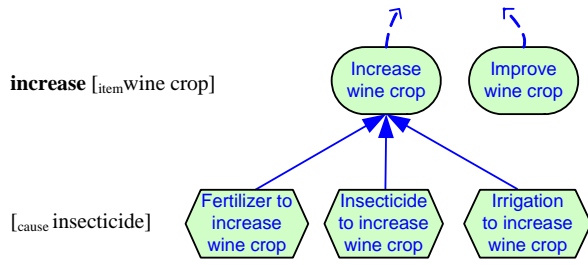
Contribution Links

Excerpt of 11 elements [Yu 1995]

II) Is Intentional Query Expansion Feasible? A Proof of Concept

Goals of other users: [increase wine crop.](#)
[improve wine crop.](#)

Goals of other users: [fertilizer to increase wine crop.](#) [insecticides to improve wine crop.](#)
[irrigation to improve wine crop.](#)



A Mockup of Utilizing Collaborative Intentional Structures to Facilitate Search on the Web

How can goals direct search behaviour? Experimentation

- Let's „buy a car“!
- Experimental Demonstrator:
<http://webdev.know-center.tugraz.at:8080/IQE3/>
[joint work with students: A. Haselsberger, C. Ruggenthaler, M. Rettenbacher]

An Exploratory Approach:

- Contacting several online resources just-in-time, such as *Yahoo Related Queries* and *43things.com Goals and Tags*
- Mining goals from identified text via POS analysis
- Structuring resulting set of goals
- Suggesting goals through **query expansions**



Project Searchengine: WT 07/08

buy a car

buy car buy cars improve credit score sell car make money get shape get license get job
find job f-k car complete education clean credit buy ?car buy kia buy enterprise buy days
buy condominium unit buy car online

Search finished in 19017 ms

Current Page: 1

Previous [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] Next

Total Search Results: 57100000

[Buy A Car Wash | Target.com](#)

Shop, find, buy "Buy A Car Wash" and other related products at Target.com.
<http://www.target.com/gp/search.html?field-keywords=buy-a-car-wash>

[Buy Cars Online, New & Used Car Search, Auto Classifieds](#)

New car prices, used cars for sale & car buying tips. ... New-Car Buying Guides. Compare Cars Side-by-Side. Compare Car Photos. Latest Rebates & Incentives ...
<http://www.cars.com/go/buyIndex.jsp>

[Buy a New Car on Yahoo! Autos](#)

Get FREE price quotes on new cars from dealers near you. ... This service is totally FREE, with no obligation to buy. Select a Car and Enter Your Zip Code ...
<http://autos.yahoo.com/newcars/buy.html?refsrc=autos/home>

[Buying a Car](#)

... buy a used car from a private party directly, or a new or used car from a car dealer. ... Cars bought from an auto dealer already have inspection stickers ...
http://www.rpi.edu/dept/doso/ISSS/public_html/carbuying.html

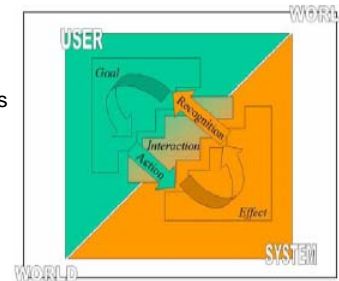
[Buying A Car Seat - Shopzilla.com](#)

Intention Recall and Precision based on Cyclic Interaction Theory

[Monk 1998]. <http://tur-www1.massey.ac.nz/~hryu/Uanalysis.pdf>

Goal-Effect Problems

- **Missing cues for goal construction**
 - Problem: Effect does not suggest appropriate goals
- **Misleading cues for goal construction**
 - Problem: Effect suggests irrelevant goals
- **Missing cues for goal elimination**
 - Problem: Effect does not eliminate completed goals
- **Misleading cues for goal elimination**
 - Problem: Effect eliminates incomplete goals



Selected Research Questions

1. How can we **identify goals** in search query logs?
2. How can we represent search goals in a **semi-formal goal graph**?
3. How can we represent the search process as a **traverse** through the goal graph?
4. How can goals **direct search behaviour**?
5. Is there a difference between **explicit and implicit intentional search queries**, and how can it be identified?
6. How can we construct **large scale goal graphs**?

Result of one of the projects conducted by
„Web-Science and Web-Technology“ students
P. Prettenhofer and A. UsSaed

- The Database of Intentions - [Battelle 2003]

“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.



Source: <http://battellemedia.com/archives/000063.php>, last accessed on July 3rd, 2007

BUT: The degree to which the goals and intentions of users can be reconstructed **varies considerable**

„Car“ → **Lower** degree of intentional explicitness
 „get loan to buy a used car in Miami“ → **Higher** degree of intentional explicitness

An intentional artifact is an electronic artifact produced by users or user behaviour that contains recognizable “traces of intent”, i.e. traces of users’ goals and intentions.

Research Overview

Research Questions:

- **How can we distinguish and identify different degrees of intentional explicitness in search queries?**

Following an exploratory, qualitative research style, we aimed to:

- Inspect a large search query log
- Analyze the way goals are expressed in search queries
- Identify possible distinction criteria
- Study differences between implicit and explicit intentional queries

Expected results:

- Increased empirically-coupled knowledge about user goals in search queries

An Exploratory Study

Data Sources:

The **AOL Search Dataset** [Pass 2006]

- ~20 mio search queries
- 657,426 unique user IDs
- *{UserID, query, timestamp, (ItemRank, URL)*}*
- Collected between March 1, 2006 and May 31 2006 by AOL

Explicit vs. Implicit Intentional Queries

Explicit intentional queries:

- Related to a specific goal in a recognizable, unambiguous way. Recognizable refers to what [Kirsh 1990] defines as "trivial to identify" by a subject within a given attention span.
- On a more practical level, this idealized definition is related to "*better queries*", or queries that have "*more precise goals*" (R. Baeza-Yates at the "Future of Web Search" workshop 2006, Barcelona).

Implicit intentional queries:

- A query, where it is difficult or extremely hard to elicit some specific goal from the intentional artifact.
- Examples include blank queries, or queries such as "car" or "travel"

Why is this distinction important?

- Disambiguation
- switching between explicit and implicit intentional queries
- Narrowing the cognitive gap

Different degrees of explicitness in search queries

- Search queries exhibit considerable variety with respect to degree of explicitness

car, 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

- Explicit vs. Implicit intentional queries

car | 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



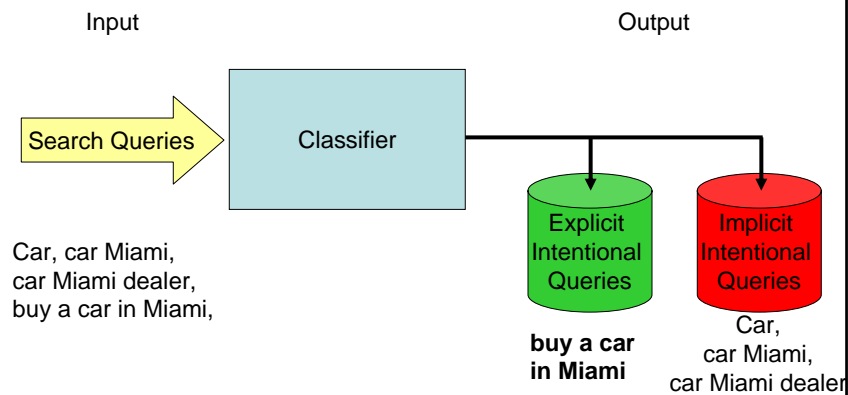
- Arbitrary selection criteria:
A query is an explicit intentional query if

- The query contains at least one verb
- The query corresponds to our definition of goals

Definition Goal: "a condition or state of affairs in the world that some agent would like to achieve or avoid"

Different degrees of intentional explicitness – Studying user goals in a large search query log

Approach: Binary classification of search queries



An Experimental Classification Approach 1/2

Part-of-Speech tagging queries:

- Only queries with query length >2 (removes ~ 60% of queries)
- Stochastic approach
- Bigram tagger (provided by NLTK)
- Tag set: Penn Treebank with 45 word classes
- Vocabulary size: approximately 13,500
- Suffix tagger as a backoff strategy (also provided by NLTK, in case of unknown words)

Tag	Description	Example
NN	Noun, sing. or mass	car
VB	Verb, base form	eat
VBG	Verb, gerund	eating
VBZ	Verb, 3sg pres	eats
JJ	Adjective	yellow
WRB	Wh-adverb	how, where
TO	"to"	to

Table 1. A sample of Penn Treebank tags (from [14])

An Experimental Classification Approach 2/2

Supervised Learning of Goal Features:

Finding $f(x) = y$: in this case, a function from $x = \text{queries}$ to $y = \text{classes}$ [explicit/implicit intentional queries]

- Part-of-speech trigrams (instead of word n-grams) to transform queries into a feature vector space

"buying/VBG a/DT car/NN"

↓
\$ \$ VBG DT NN \$ \$ → \$ \$ VBG, \$ VBG DT, VBG DT NN, DT NN \$, NN \$ \$

- Random sample from the set of queries which contain at least one verb (1,598,612 out of 20,494,002 queries)
- Training set: 98 instances, 59 positives, 39 negatives
- Naive bayesian classifier (provided by Weka) trained on the feature vectors using 10-fold cross-validation
- 20 most predictive features (experimental exploration)

\$ \$ NN	\$ \$ VBG
\$ WRB TO	WRB TO VB
\$ NN NN	\$ VBG DT
VBG DT NN	\$ VBG NN
\$ \$ VBZ	JJ NN \$
\$ VBG IN	VBG IN NN
\$ VB NN	TO VB VBN

Results I

	Entire Dataset	Condensed Dataset
Queries	20,494,002	279,260
Explicit Intentional Queries	346,349-616,869	138,513-163,089
Implicit Intentional Queries	19,877,133-20,147,653	116,172-140,747
Explicit Intentional Queries, 95% confidence interval	1.69% - 3.01%	49.6% - 58.4%
Users	657,426	94,487

Table 3. Statistical overview of the condensed dataset

Are there interesting differences between these two datasets?

What kind of

rectly / incorrectly?

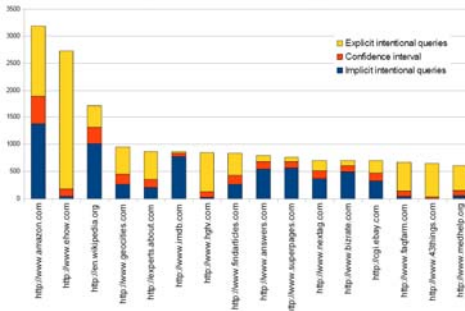


Figure 3. Top 16 websites in the condensed dataset

True positives

Correctly Classified Intentional Queries
"buying groceries online"
"how to get revenge on neighbor within limits of law"
"helping children handle death of a loved one"
"cleaning the ak-47"
"coughing up blood"
"dealing with the guilt of cheating"

Table 4. Examples of correctly classified queries

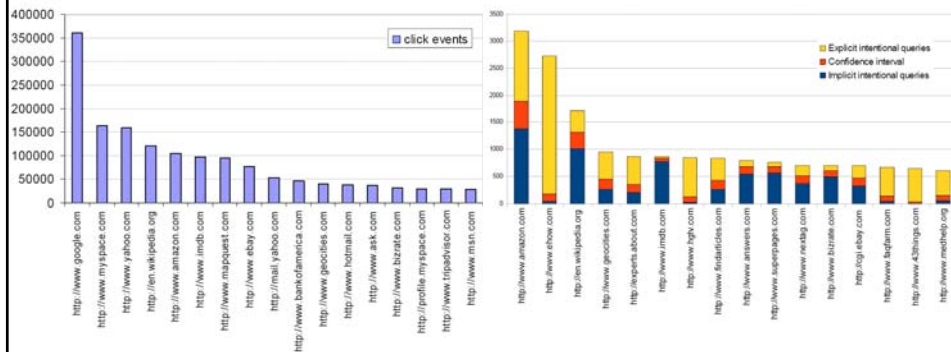
False positives

Incorrectly Classified Intentional Queries
"saving privat ryan"
"driving school illinois"
"stem cell transplant"
"founding fathers temple"
"recovering the satellites lyrics"

Table 5. Examples of incorrectly classified queries

Results III

Do more explicit intentional queries yield different click-through result sets?



Top 16 websites in the entire dataset

low proportion of explicit intentional queries
(1,69-3,01%)



Top 16 websites in the condensed dataset

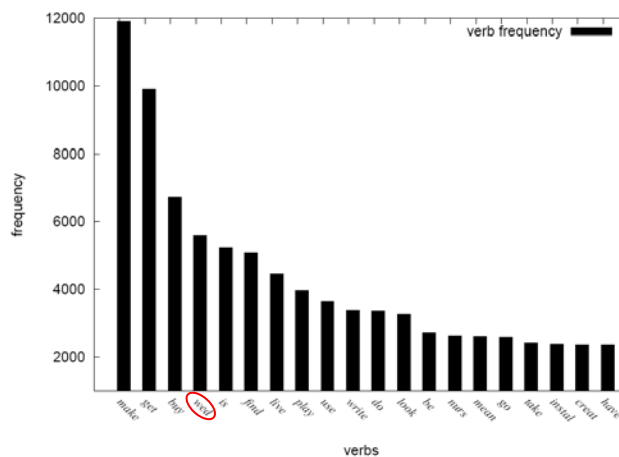
high proportion of explicit intentional queries
(49,6-58,4%)

Experimental Classification

Examples: 43things.com (from #388 up to rank #15), ehow.com (from #64 up to #2), hgtv.com (from #97 up to #7), and medhelp.org (from #104 up to #16).

Results IV

What are verbs indicating the presence of goals in queries?



Verb frequency histogram (condensed dataset)

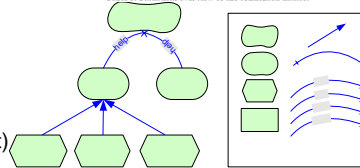
Next Steps

In the medium term, ...

- We intend to develop methods and techniques that support
 - *Intentional Classification of Artifacts and*
 - *Extraction of Goals from Intentional Artifacts and Social Corpora*
- to construct large-scale
 - *Goal Association Graphs* (X is associated with Y) and
 - *Intentional Networks* (different intentional relations and nodes, see image to the right)
- that enable novel applications, such as
 - *Intentional Query Expansion*
 - *Intentional Metadata*
 - *Intentional Analysis*

	Entire Dataset	Condensed Dataset
Queries	20,494,002	279,260
Explicit Intentional Queries	346,349-616,869	138,513-163,089
Implicit Intentional Queries	19,877,133-20,147,653	116,172-140,747
Explicit Intentional Queries, 95% confidence interval	1.69% - 3.01%	49.6% - 58.4%
Users	657,426	94,487

Table 3. Statistical overview of the condensed dataset



Idea 1

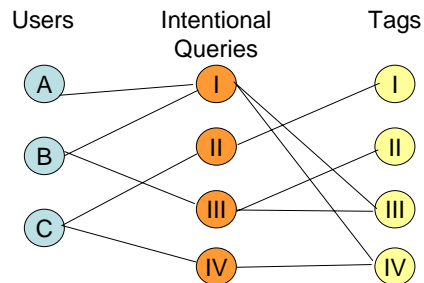
- We have two unequally distributed types of queries
- Explicit and implicit intentional queries
 - Distinction: Explicit intentional queries contain verbs and thereby allow to more easily grasp the intentionality behind a query than implicit intentional queries
 - Example: „car shop“ vs. „buy car“, „sell car“, „repair car“

	walmart [1]	2006-04-29 22:06:40	http://www.walmart.com	1
implicit intentional queries	walmart [1]	2006-04-29 22:06:40	http://www.walmart.com	1
	myspace.com [1]	2006-05-01 14:20:56	http://www.myspace.com	1
	target [1]	2006-05-01 14:26:25	http://www.target.com	1
Explicit intentional queries	how to get rich breast milk [1]	2006-05-02 01:21:08	http://www.nlm.nih.gov	2
	breast milk [1]	2006-05-02 01:28:34	http://www.babycentre.co.uk	2
Explicit intentional queries	how to have good breast milk [1]	2006-05-02 01:32:34	http://www.breastfeeding.com	1
	how to have good breast milk [1]	2006-05-02 01:32:34		0
	yellow breast milk [1]	2006-05-02 01:40:47		0
	yellow breast milk [1]	2006-05-02 01:40:56	http://www.dummies.com	1
	yellow breast milk [1]	2006-05-02 01:40:56	http://www.mother-2-mother.com	2
	breast feeding and going back to work [1]	2006-05-02 01:50:55		0
implicit intentional queries	breast feeding and going back to work [1]	2006-05-02 01:50:58	http://www.askbaby.com	1
	breast feeding and going back to work [1]	2006-05-02 01:50:58	http://www.growingwe	3
	breast feeding and going back to work [1]	2006-05-02 01:50:58	http://www.meds	2
	nestle formula [1]	2006-05-02 02:01:36		0
	free nestle formula [1]	2006-05-02 02:01:57	http://www.kidneedies.com	1
	free nestle formula [1]	2006-05-02 02:01:57	http://www.epinio	1
	free nestle formula [1]	2006-05-02 02:01:57	http://www.verybestbaby.com	2
	free nestle formula [1]	2006-05-02 02:01:57	http://www.bri...beginnings.com	6
	good start [1]	2006-05-02 02:09:33	http://www.verybestbaby.com	1
	what fenugreek [1]	2006-05-02 02:11:07	http://www.breastfeeding.com	1

How can we automatically construct a network model based on this information?

Network Analysis

- Analyzing the tripartite graph of Search
 - Consisting of users, explicit intentional queries and tags



Based on this conceptualization, the following two-mode networks can be folded into one mode networks:

- Intentional Queries – Tags
- Users – Intentional Queries
- Users - Tags

Intentional Metadata: Research Opportunities

- Intentional XHTML
 - `label` link relationships with `rel="..."` 'rel' is extensible and multivalued `rel="relationship1 relationship2"`
- Intentional Microformats
- Intentional (Search?)Browser-Plugins
- Intentional Tags / Geotags
- Intentional Webservice Descriptions
 - e.g. WSMO including `wg` mediators (web service – goal) + `gg` mediators (goal -goal)
- Automatic Intentional Metadata Creation (Corpus-based)
- Intentional Weblog Tags

- → SPECIFIC EXAMPLES

Administrative Issues

Final Exam in two weeks

- No repetition of previous home assignments
- Understanding of basic network constructs (e.g. centrality, weak ties, formal models of small world networks, etc)
- Scope: All lecture slides and mandatory readings
- **You need to register for the exam with TUG online**
- Date: 22.1. 14:00, HS Modul
- Results: within 4 weeks
- „Einsichtnahme“: Fr, 29.2. 2008, 10:00 – 11:00

Next week: Guest lecture (attendance list will be handed out)

- **Web Technologies II, Peter Scheir, KMI**
The semantic web represents a current research effort to increase the capability of machines to make sense of content on the web. In this class, Peter Scheir will give a guest lecture on the basic principles underlying the semantic web vision, including RDF, OWL and other standards.

Any questions?

See you next week!