

Functional Faceted Web Query Classification

Bang V. NGUYEN and **Min-Yen KAN**

Department of Computer Science
School of Computing
National University of Singapore



Web queries and their analysis

aloha airlines

color blindness

kazaa lite

house document
no. 587

What can we see?

Different types of queries!

1. Navigational

2. Informational

2.1 Directed

2.1.1 Closed

2.1.2 Open

2.2 Undirected

2.3 Advice

2.4 Locate

2.5 List

3. Resource

3.1 Download

3.2 Entertainment

3.3 Interact

3.4 Obtain

Rose and
Levinson,
2004

Related Work

- **Theory:** earlier Classifications (Broder, 2000; Rose and Levinson, 2004)
 - Originally to uncover latent user intent
 - Largely mutually exclusive
- **Practice:** Automated classification
 - Kang and Kim (2003)
 - Lee et al. (2004)
 - and others ...

A clear mismatch between theory and practice

Functional Faceted Classification

- **Functional**: meaning that an automated engine can use the results to act appropriately
- **Faceted**: meaning that queries can manifest different, fine grained behavior in combination
 - Mostly independent facets

Outline

- How to formulate such a classification scheme?
 1. How does such a classification relate to previous work?
 2. How are web queries distributed given this scheme?
 3. How well do human subjects classify against this scheme?
- Just briefly: how to automate such a classification?

A Faceted Classification Scheme

Goals of the classification:

1. Should have high coverage
2. Have discrete values
 - For ease of replication and analysis
3. **Facets** should be largely independent
 1. Different combinations of facets possible
4. Classification leads to **action**

Went through several iterations, settled on four facets

- Others considered but dropped

Query: Stanford University

1. Ambiguity = **Specific**
2. Authority sensitive? = **Yes**
3. Temporally sensitive? = **No**
4. Spatially sensitive? = **No**

Most facets have been proposed before by others
We integrate them into one classification scheme
Let's look at each facet in more detail

Ambiguity

- Modeled after a library agent's interpretation of ambiguity (Stojanovic, 2005)
- Actionable strategies:
 - Disambiguation / Categorization
 - Suggest modifiers for **general** queries

Examples:

- **Polysemous**: contain many senses
 - “mustang”
- **General**: cover multiple sub-categories
 - “health”, “travel”
- **Specific**: address a coherent set of relevant information
 - “apple iphone”

Authority Sensitivity

Does the query refer to a well-known concept or an authoritative answer?

- Actionable strategies:
 - Jump directly to or extract authoritative results
 - Favor prestige factors (e.g. PageRank) in ranking results

Examples

- Yes

- “hsbc internet banking”
- “U.S. independence day”

- No

- “laptop harddrives”
- “science supplies”

- New facet, not previously mentioned by literature
 - Related to trust in question answering (Lin *et al.*, 2003)

Temporal and Spatial Sensitivity

Should the query results **change** with respect to the user's temporal or spatial **context**?

- Lots of related work; just a sample:
 - Categorizing web queries according to geographical locality (Gravano *et al*, 2003)
 - Information diffusion through blogspace (Gruhl *et al*, WWW '04)
 - Temporal relation between queries (Chien *et al*, WWW '05)
- Actionable strategies:
 - Suggest temporal or spatial modifiers
 - Make users aware of spatially / temporally changing
 - Take locality of users into account

Temporal Examples

- **Yes**
 - “university ranking”, “u.s. president”
- **No**
 - “Toyota Canada”, “gravity forces”

Spatial Examples

- **Yes**
 - “transport services”, “pizza order”
- **No**
 - “email signature”, “how to run fast”

Outline: Research questions

- How to formulate such a classification scheme?
 1. How does such a classification relate to previous work?
 2. How are web queries distributed given this scheme?
 3. How well do human subjects classify against this scheme?
- Just briefly: how to automate such a classification?

1. Relate to R&L's taxonomy

<Ambiguity, Authority, Temporal, Spatial>



1. Navigational	<Specific, Y, N, N>
2. Informational	
2.1 Directed	
2.1.1 Closed	<Specific, Y, -, ->
2.1.2 Open	<Specific, N, -, ->
2.2 Undirected	<-, N, -, ->
2.3 Advice	<Specific, -, -, ->
2.4 Locate	<Specific, Y, -, Y>
2.5 List	<Specific, -, -, ->
3. Resource	
3.1 Download	<Specific, Y, Y, N>
3.2 Entertainment	<-, -, -, ->
3.3 Interact	<Specific, Y, -, ->
3.4 Obtain	<Specific, Y, -, ->

Observations:

- Previous work only dealt with specific queries
- Previous work often not actionably differentiated – have same facet signature

- Bottom Line: not very strongly correlated – need actionable classification

2. Actual Query Distribution

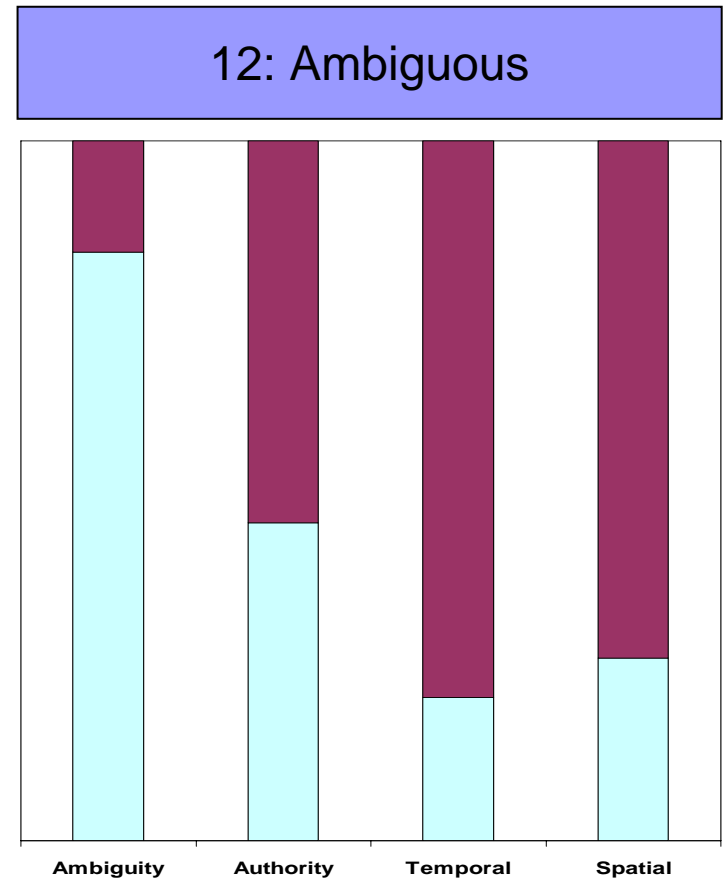
- Manual classification
 - 100 queries (limited, small indicative sample)
 - Randomly chosen
 - Reject non-English, offensive queries
 - From AllTheWeb™, 2002
- Judged by authors
 - Use only the query string and search results as evidence
 - Other data (e.g., clickthrough data) intentionally left out
 - Broader impact

12: Ambiguous

88:
Non ambiguous:
General and
Specific

Observations

- *Specific* queries make up a majority.
- Almost half of require authoritativeness
- Temporally and spatially sensitive queries amount to about 1/5th.
 - Percentage growing with mobile devices?



3. Replicability by human subjects

Goals:

- A. Is our classification replicatable / understandable?
- B. Are the coarse granularity of facet values ok?

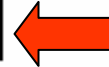
- Subsampled 75 of the 100 queries
- Asked 25 human volunteers to judge queries
- Each query is judged by 5 evaluators.
 - Capture the average rating of all users on each facet on Likert 5-point scale

(1 sensitive, 5 insensitive)

(1 specific, 5 ambiguous)

Results: ANOVA and polarity

Facet	F-ratio	P-value
<i>Ambiguity</i>	0.4297	0.7871
<i>Temporal</i>	1.9353	0.1039
<i>Spatial</i>	0.7223	0.5770
<i>Authority</i>	4.7070	0.0010



Facet	# Bipolar	# Non-bipolar
<i>Ambiguity</i>	63	4
<i>Authority</i>	57	10
<i>Temporal</i>	64	3
<i>Spatial</i>	65	2

ANOVA: do people agree?

- P-value > .01 people do not differ significantly

Conclude:

- Ambiguity, Temporal and Spatial Sensitivity OK
- Authority needs work, not consistent

Bipolar (1 or 5) vs.
Non-bipolar (2, 3 or 4)

Conclude:

- Subjects satisfied with coarse grained values
- Validates our value choices

Outline: Research questions

- How to formulate such a classification scheme?
 1. How does such a classification relate to previous work?
 2. How are web queries distributed given this scheme?
 3. How well do human subjects classify against this scheme?
- Just briefly: how to automate such a classification?

Automatic classification

Most facets already have prior work for automated classification

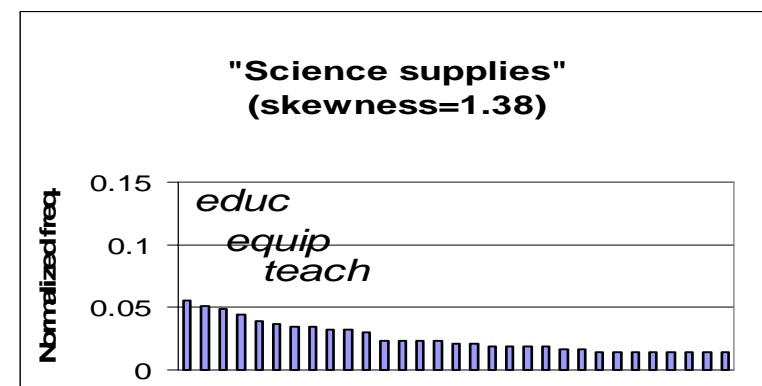
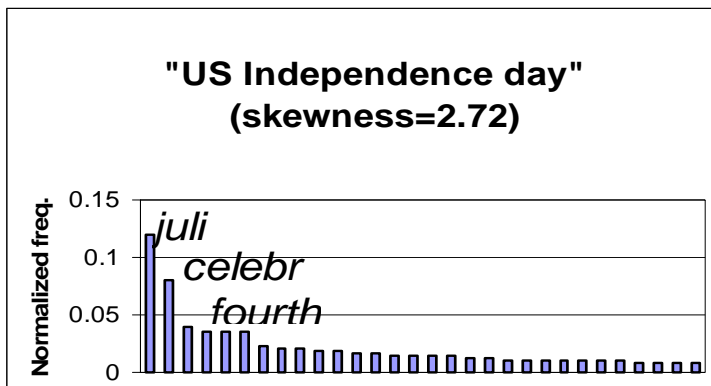
- Ambiguity:
 - Analyzing content topical distribution; lexical databases
- Temporal and spatial sensitivity: mining of query log to find temporal modifiers
 - “U.S. president” and “U.S. president 1966
 - measure changing results via content publishing (e.g. RSS)

However, **authority sensitivity** has not yet been explored

Auto Authority Sensitivity Classification

- Hypothesis: authoritative queries often have answers repeated in search results
- Idea: Look for repeating information (c.f., trusted question answering)
 - Examine summary snippets for query
 - Extract keywords and measure their distribution
 - Capture keyword histogram's *skewness*:

$$skewness = \frac{\sum_{i=1}^M \left(\frac{(freq(i) - \mu)}{\sigma} \right)^3}{M}$$



Preliminary Results

- Use just this one feature
- Learn a classification boundary
 - SVM^{light}
- 78%, 60%, 68% for Precision, Recall, F1

Conclude (Indicative)

- Automatic classification is possible for authority sensitivity

Conclusions

- Proposed a faceted actionable scheme with four facets:
 - *Ambiguity, Authority, Temporal and Spatial Sensitivity*
- Related our classification to previous work
- Showed reliability: *authority* facet perhaps not well defined
- Gave indicative distribution based on small query set

- Future work:
 - Implement a fully automated classification system
 - Larger scale evaluation and distributional analysis
 - Investigate possible expansion of the taxonomy
 - Requirement for special collection?