gesis
Leibniz Institute for the Social Sciences

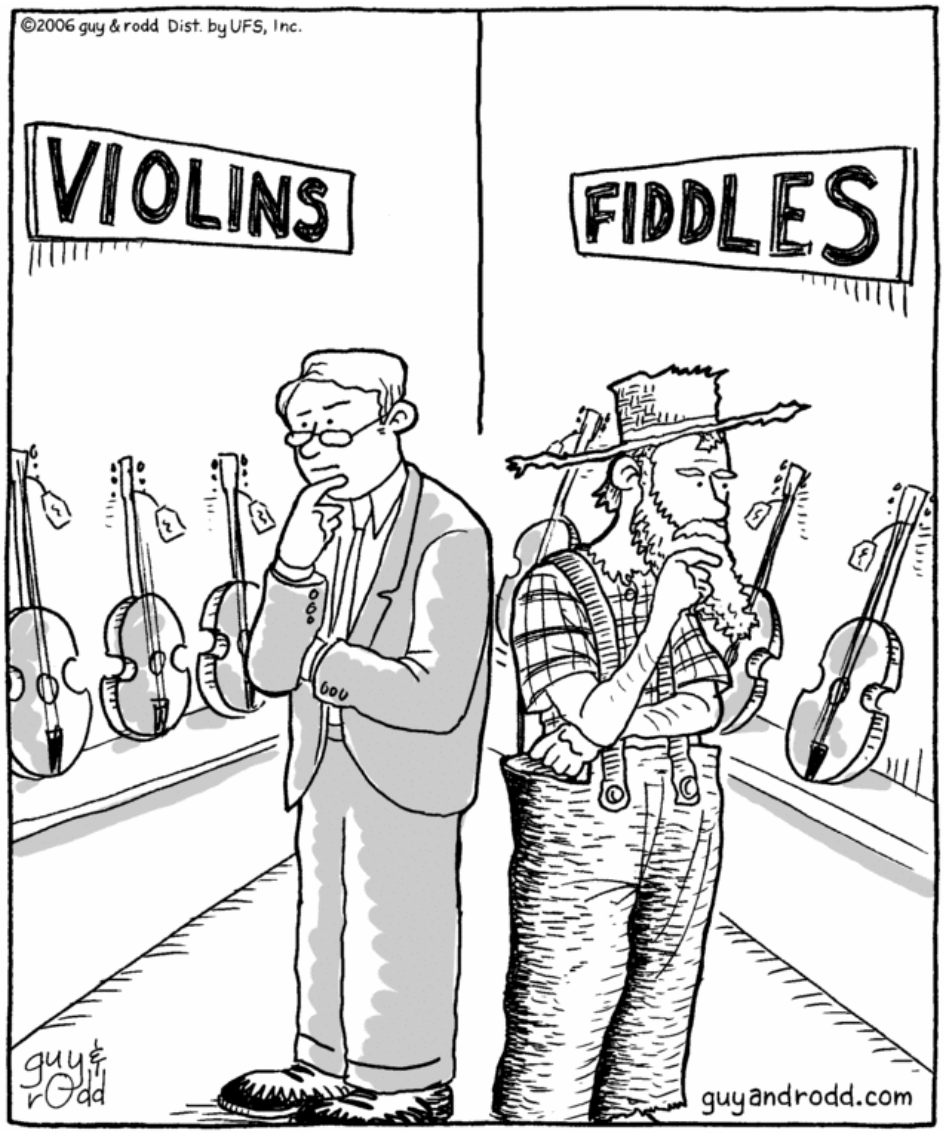
User Interface Design for Search Term Recommendation and Interactive Query Expansion Services

8th NKOS Workshop at 13th ECDL 2009

Philipp Schaer
with the help of Daniela Holl and Dominik Hering
philipp.schaer@gesis.org
Corfu, 2009-10-01

Overview

Motivation	• The word/vocabulary problem
Background	• Co-Word Analysis with LSA, pLSA, SVM • Human Computer Inform. Retrieval
Prototyp	• Mindserver, technical basis
User Interface	• Pre- and Post-Search approach • Own pretests and conclusions
Outview	•Future work


©2006 guy & rodd Dist. by UFS, Inc.
VIOLINS
FIDDLES
guy & rodd
guyandrodd.com

The Vocabulary Problem

- Coined by Furnas et al. (1987)
- Especially a problem in Digital Libraries

„I choose search terms based not specifically on the information I want, but rather on how I could imagine someone wording [...] that information."

from Aula et al. (2005)

The Berkley Search Term Recommender

Ohsumed Specialty Search Term Recommenders

cardiac arrest [Clear]

Geriatrics [STR One] [STR Two] [STR Three] Suggested Mesh Headings 10

Anesthesiology Specialty

Query: cardiac arrest

1 Heart Arrest
2 Heart Surgery
3 Cardiac Output
4 Respiratory Insufficiency
5 Heart Arrest, Induced
6 Heart
7 Heart Diseases
8 Resuscitation
9 Coronary Artery Bypass
10 Hyperkalemia

Drug Therapy Specialty

Query: cardiac arrest

1 Heart
2 Purkinje Fibers
3 Myocardium
4 Anti-Arrhythmia Agents
5 Arrhythmia
6 Heart Conduction System
7 Cardiac Output
8 Myocardial Contraction
9 Anilides
10 Heart Arrest

Geriatrics Specialty

Query: cardiac arrest

1 Coronary Disease
2 Heart Diseases
3 Crime
4 Heart
5 Cardiovascular Agents
6 Mitral Valve Insufficiency
7 Cardiomyopathy, Hypertrophic
8 Aortic Valve Insufficiency
9 Up-Regulation (Physiology)
10 Pindolol

from Buckland (2006)

Co-Word analysis

d1
Title: Introduction to Information Retrieval
Controlled Terms: Information, retrieval

d2
Title: A survey of Information Retrieval and filtering methods
Controlled Terms: Information, retrieval

d3
Title: Expanding queries for the web
Controlled Terms: Information, retrieval, query expansion

d4
Title: Efficient and self-tuning incremental query expansion for top-k query processing
Controlled Terms: Query expansion

Co-word analysis by i.e.
LSA, pLSA, SVM, ...

Co-Word analysis with Mindserver

- Commercial full text classifier (i.e. used by Reuters)
- Probabilistic Latent Semantic Analysis (pLSA) and Support Vector Machines (SVM)
- Training sets:
  - Social science database SOLIS (370.000 documents, title, abstract and controlled thesaurus terms)
  - Others: CSA-SA, CSA-PEI, SPOLIT, FES, …
- Alternatives
  - Lextek, MALLET, Nvivo, Readware, …

Our first prototype on basis of Sowiport

from Holl (2009)

from Holl (2009)

One HCIR principle: Analyse user requirements

Berkley	• No integration and coupling into a search environment
Sowiport	• Integrated into a search environment • BUT: Additional step required
Prototype	• Integration • One stop precedure

Search-Term-Recommender in use

Pre-Search-Recommender

Post-Search-Recommender

Paperprototyping

Paperprototyping

- Low-fidelity Prototype on paper
- Every sheet of paper with different interactions steps
- Content was generated by Mindserver
- 5 persons (age: 28 – 63; med. – very experienced)
- A virtual help button was available

Term suggestion based on term 1+X

vs

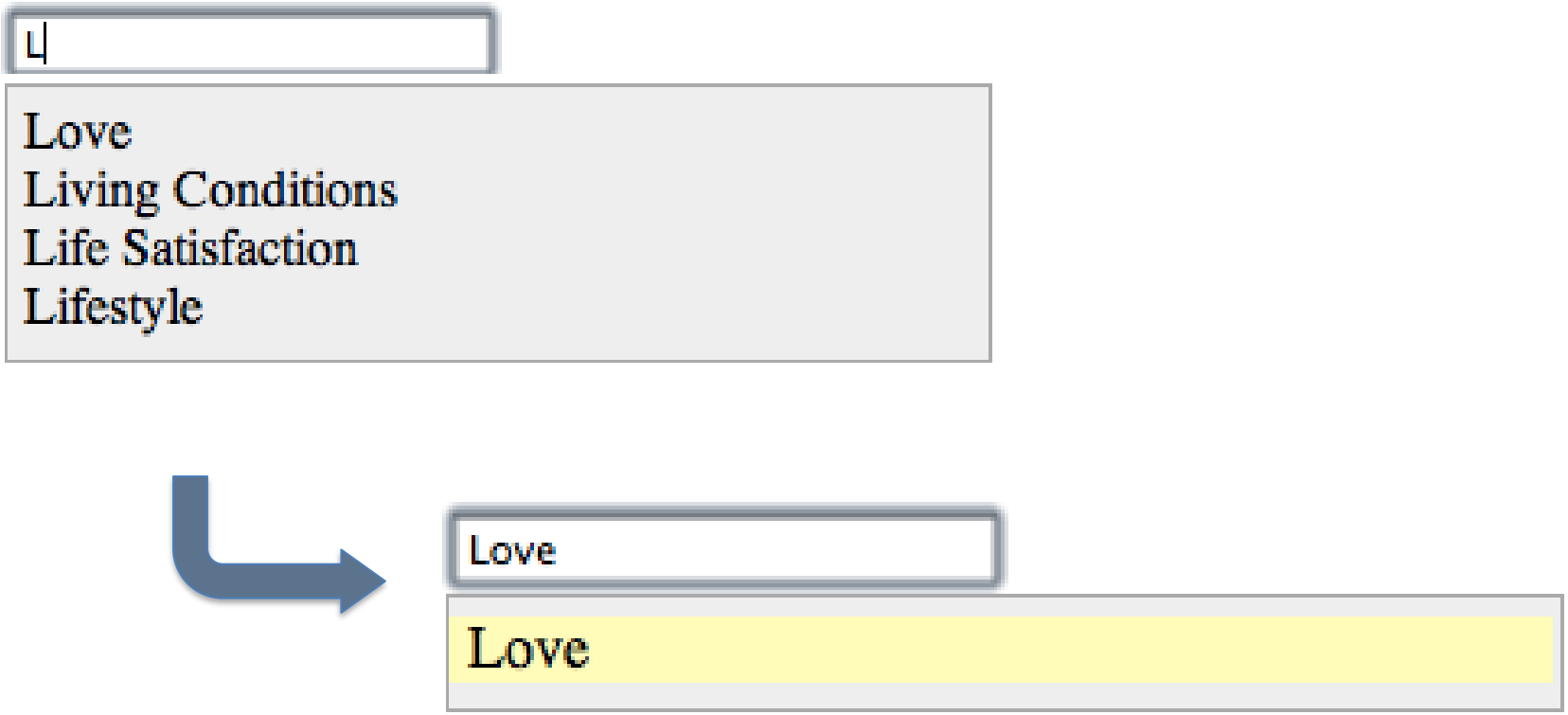
Term suggestion based on term 1, then 2, then 3,…

Paperprototyping: Findings

- Overall positive feedback
- 4 of 5 persons knew drop-down term suggestions (Ebay, Google, etc.)
- Too early suggestions lead to confusion
- 4 of 5 persons liked variante 2 more
- Experienced users recognized variations (redundancy, options, number of terms)

Fast, cheap and effective way of prototyping

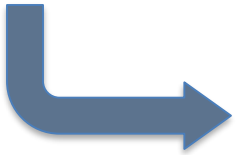
Walkthrough: Step 1 “Looking for Love”



Walkthrough: Step 2 “Love + Space”

Love + Space

Dating (Social)
Opposite Sex Relations
Sexual Intercourse
Interpersonal Relations
Sexual Behavior
Happiness
Intimacy
Family Structure



Love Money

gesis
Leibniz Institute for the Social Sciences

Walkthrough: Step 3 “Love Money”

Term: Love Money Senden

Total hits: 2261

1. Prasad, Monica : *The Morality of Market Exchange: Love, Money, and Contractual Justice* (1999)
2. Cerulo, Massimo : *For Love or Money. The Commercialization of Intimate Life* (2008)
3. Ballinger, Lee : *In your face Sports for love and moneyVor Deinen Augen Die Kommerzialisierung des Sports* (1981)
4. Onnen-Isemann, Corinna : *Money and Love. The Symbolic Significance of Money in Couple Relationships* (2004)
5. Nassehi, Armin : *Money and Love. Symbolic Significance of Money In a Couple's Relationship* (2005)
6. Stewart, B.K. : *The Regulation Of Sport Agents: For The Love Of Money* (1998)
7. Lopes Junior, Edmilson : *Love, Sex and Money: A Sociological Interpretation of the Market of Sexual Services* (2005)
8. Sawhill, Isabel; Thomas, Adam : *For Love and Money? The Impact of Family Structure on Family Income* (2005)
9. Binder, Amy : *For love and money: Organizations' creative responses to multiple environmental logics* (2007)
10. Bean, Frank D.; Van Hook, Jennifer; Brown, Susan K. : *For Love or Money? Welfare Reform and Immigrant Naturalization* (2006)

next 10 documents >>

Economic Sociology Love Money Eroticism Dating (Social) Sexual Intercourse Opposite Sex Relations Interpersonal Relations Sexual Behavior Sexual Inequality Economic Theories Marital Relations

- Economic Sociology [0.9582742]
- Public Sphere [0.42437306]
- Love [1.0]
- Value (Economics) [0.8274064]
- Individual Differences

Next steps and conclusion

Next steps

- More sofisticated evaluation
- Multi-database scenarios?
- Offer as Web-Service (see Bibsonomy)?

Conclusions

- STR is practical and evaluated procedure
- STR can be implemented using a commercial fulltext classification system
- It needs to be fitted to the users needs

http://www.gesis.org/index.php?id=2479

References

A. Aula, N. Jhaveri, and M. Käki. Information search and re-access strategies of experienced web users. Proceedings of the 14th International Conference on World Wide Web (WWW'05), pages 583–592, 2005.

Michael Buckland (2006). Collaboration: Bad Words and Strong Documents. COOP2006 7th International Conference on the Design of Cooperative Systems, Marseille, May 10, 2006

GW Furnas, TK Landauer, LM Gomez, and ST Dumais (1987). The vocabulary problem in human-system communication. Communications of the ACM, 30(11):964–971, 1987.

Dominik Hering (2009). Interaktionsdesign für Search Term Recommender. Master Thesis. Universität Koblenz-Landau. (to appear)

Daniela Holl (2009). Search Term Recommender auf Basis der Software MindServer. Master Thesis. Universität Koblenz-Landau.

Vivien Petras (2006). Translating Dialects in Search: Mapping between Specialized Languages of Discourse and Documentary Languages. PhD thesis, University of California, Berkeley, 2006.

Gary Marchionini, Ryen White, Nick Belkin, Gene Golovchinsky, Diane Kelly, Peter Pirolli, mc schraefel (2008). Information Seeking Support Systems: An invitational workshop sponsored by the National Science Foundation, http://ils.unc.edu/ISSS/