June 2 Presentation of Group Project

1. 3L党
2. lemon group
3. 中关村都市丽人
4. 你吃对了吗？
5. 学术幼儿园
6. 张奕桢先生
7. 心琦小组
8. 熟悉的异乡人
9. 陈氏姐妹

June 2 Presentation of Group Project

- 每组汇报时长为10分钟，问答时间5分钟

① 课前助发放评分表，对除自己组之外的其它小组打分和评价；下课时提交，作为期末小组报告得分的重要依据之一；

② 各小组严格控制汇报时间，有专人记录，超时会被打断，并影响得分；

③ 每个小组需要对除自己组之外的至少1个小组的展示提出问题。

June. 9 Presentation of Group Project

(1) Research Background and specific research questions，30%
包括研究背景、明确且有逻辑的研究问题。

(2) Research method，30%：
包括研究方法选择、设计和实施过程。

(3) Results and Conclusion，30%：
包括数据分析过程，结果发现与讨论，研究不足或下一步完善研究计划。

(4) Presentation skills，10%：
包括流畅度、时间把握等。

北京大学

Term Paper (DDL: June 9, 22:00)

期末报告作业长度在10页以内

- 概述 (10分): 研究背景、动机和研究问题
- 文献回顾 (20分)
- 研究方法 (30分)
  - 对测试系统或界面功能的介绍
  - 包括研究设计、研究流程、评估指标
  - 对评估指标进行操作化，访谈或问卷问题如何与指标对应
- 测试实验结果 (30分)
- 研究不足及下一步研究设想(10分)
- 参考文献
- 附录

Term Paper (DDL: June 9, 22:00)

- Introduction (10)
  - Addresses Research background, motivation, and Research questions
- Related Work (10)
  - Uses course readings appropriately
  - Include papers that each of you have annotated
  - Includes core relevant literature
- Methods (40)
  - Selects appropriate evaluation criteria and measures
  - Research design is appropriate
  - Procedures are described in detail
  - An understanding the methods' advantages and disadvantages
- Results (20)
  - Present main data analysis and the overall results
- Limitation and Future work (10)
- References
- Appendix

Lecture 11 Relevance Feedback

刘畅

北京大学

Relevance Feedback（相关反馈）

- Take advantage of user relevance judgments in the retrieval process:
  - User issues a (short, simple) query
  - User gets back an initial hit list
  - User marks hits as relevant or non-relevant
  - The system computes a better representation of the information need based on this feedback
  - Single or multiple iterations (although little is typically gained after one iteration)

三种相关反馈方式

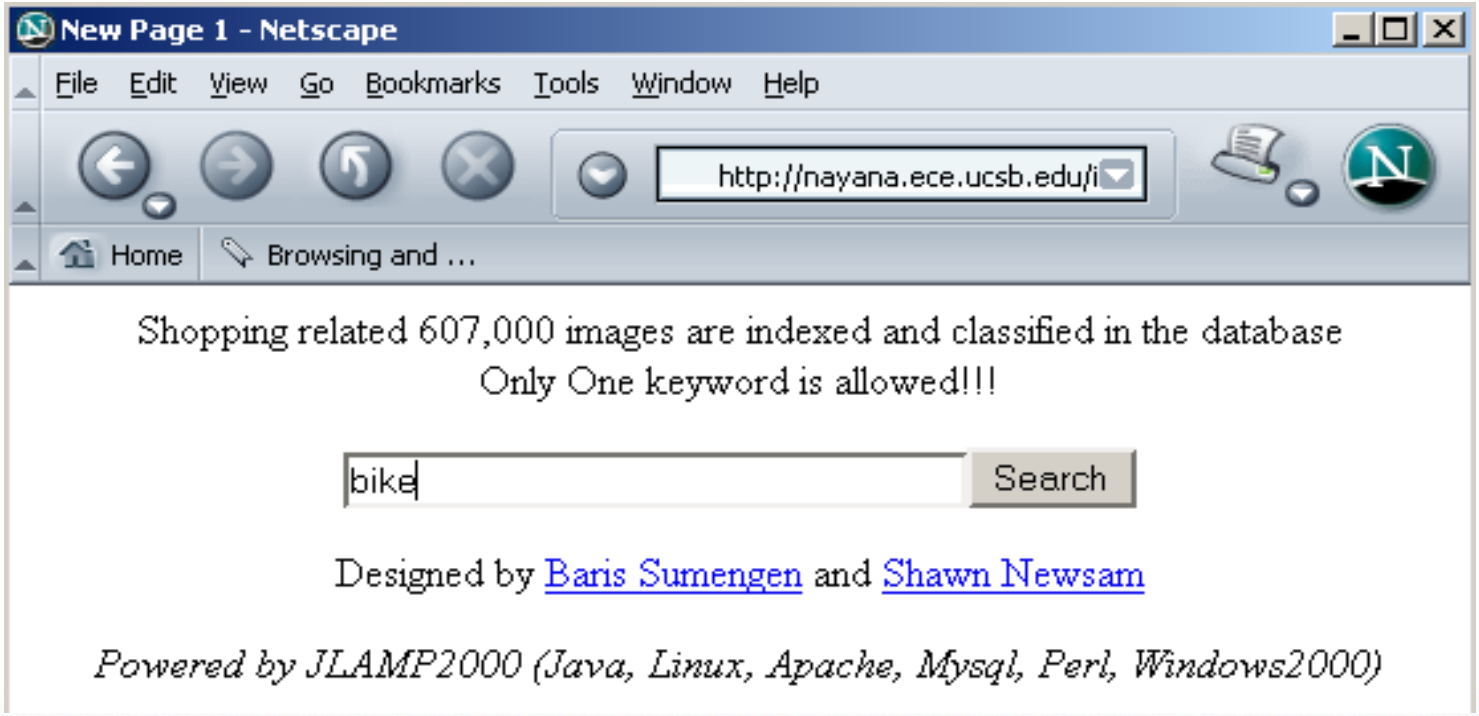
- Explicit feedback （显性相关反馈）
  - users explicitly mark relevant and irrelevant documents
- Implicit feedback （隐性相关反馈）
  - system attempts to infer user intentions based on observable behavior
- Pseudo relevance feedback（伪相关反馈）
  - feedback in absence of any evidence, explicit or otherwise

Why relevance feedback?

- Belkin's ASK model:
  - You may not know what you're looking for, but you'll know when you see it
  - Query formulation may be difficult; simplify the problem through iteration
- Facilitate vocabulary and concept discovery
- Boost recall
  - "*Find more documents like this…*"

Relevance Feedback Example

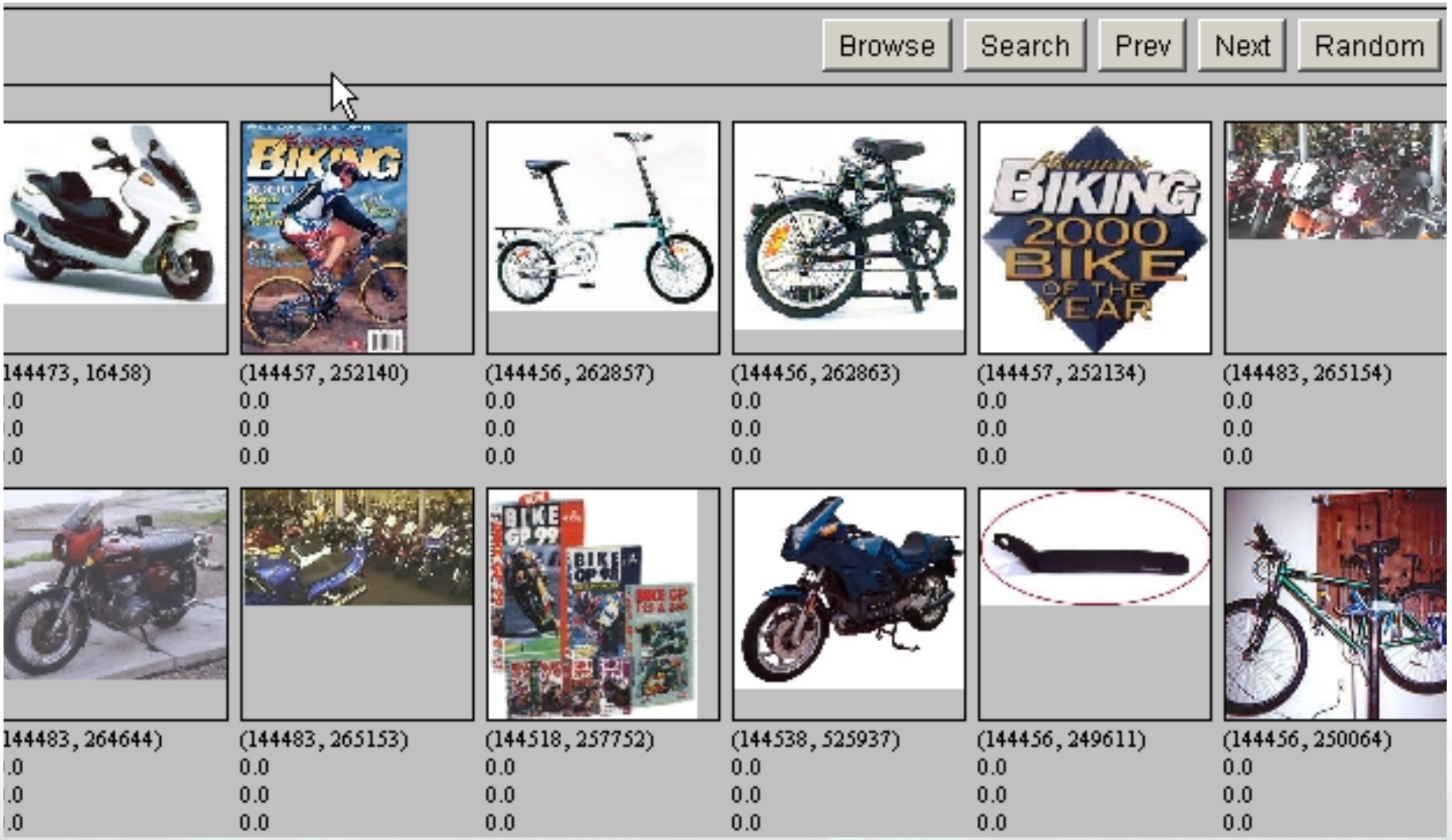
Image Search Engine



Acknowledgement：部分内容参考Jimmy Lin@University of Maryland的相关课件，在此表示感谢。

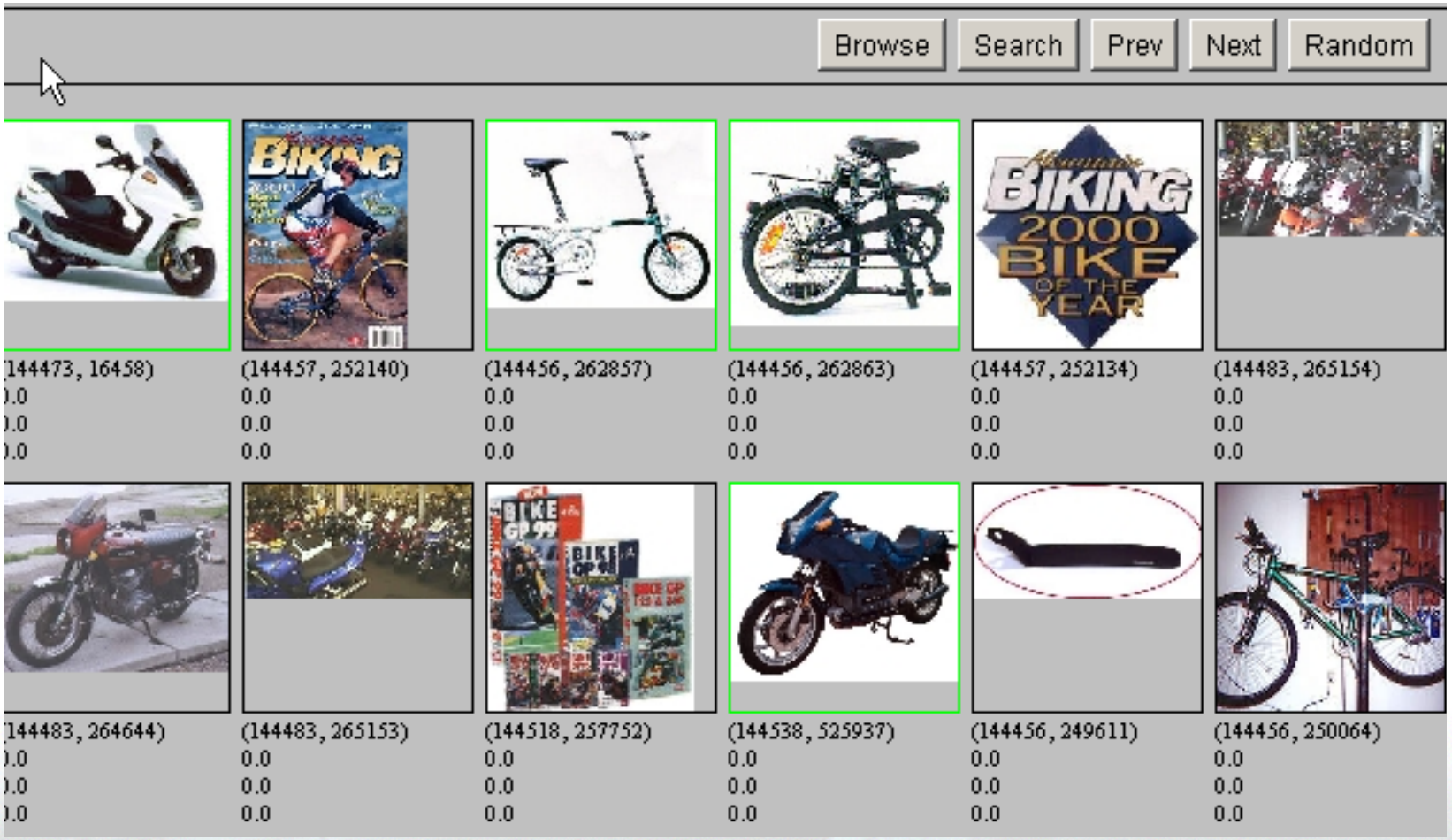
北京大学

Initial Results



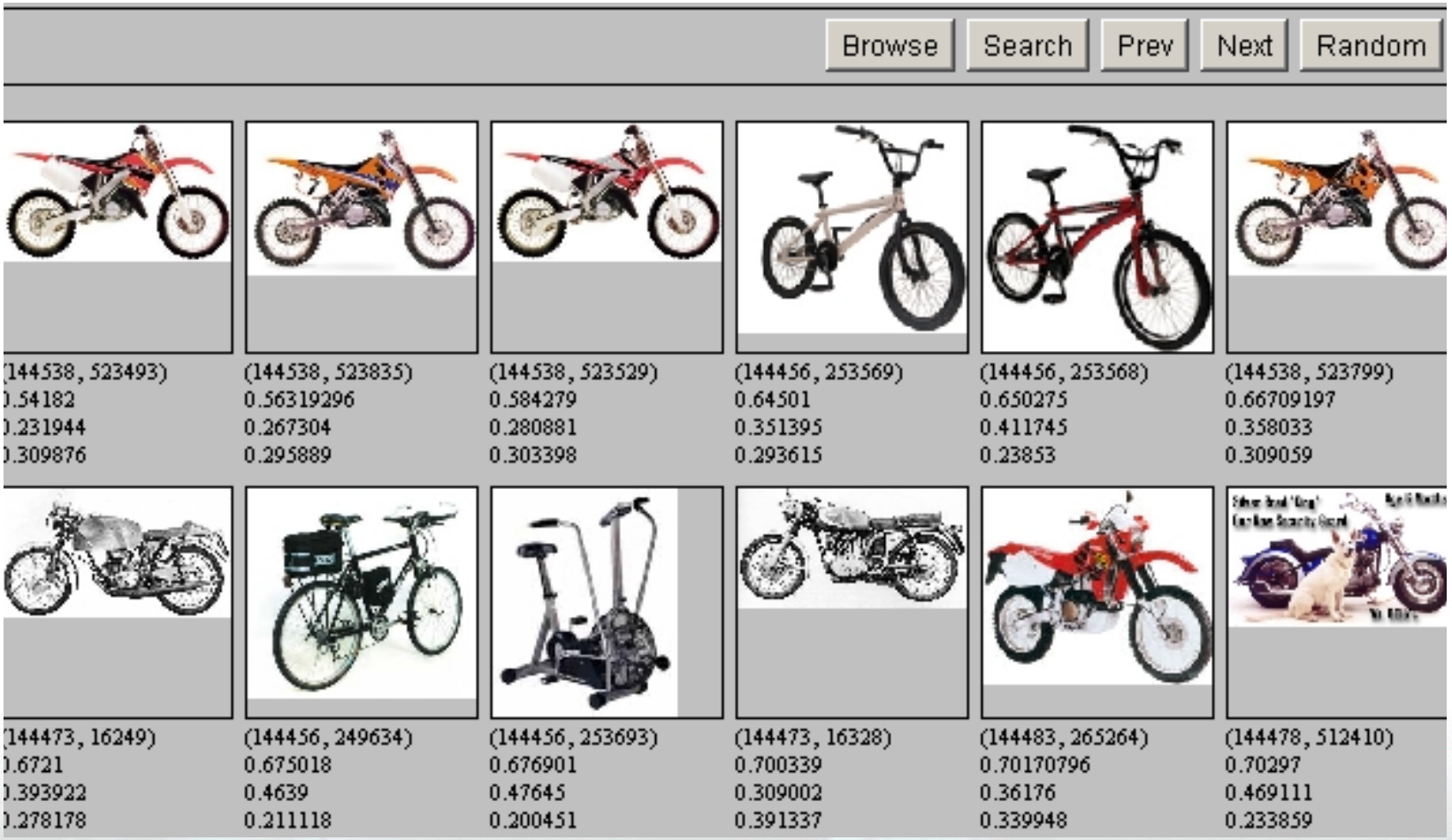
北京大学

Relevance Feedback



北京大学

Revised Results

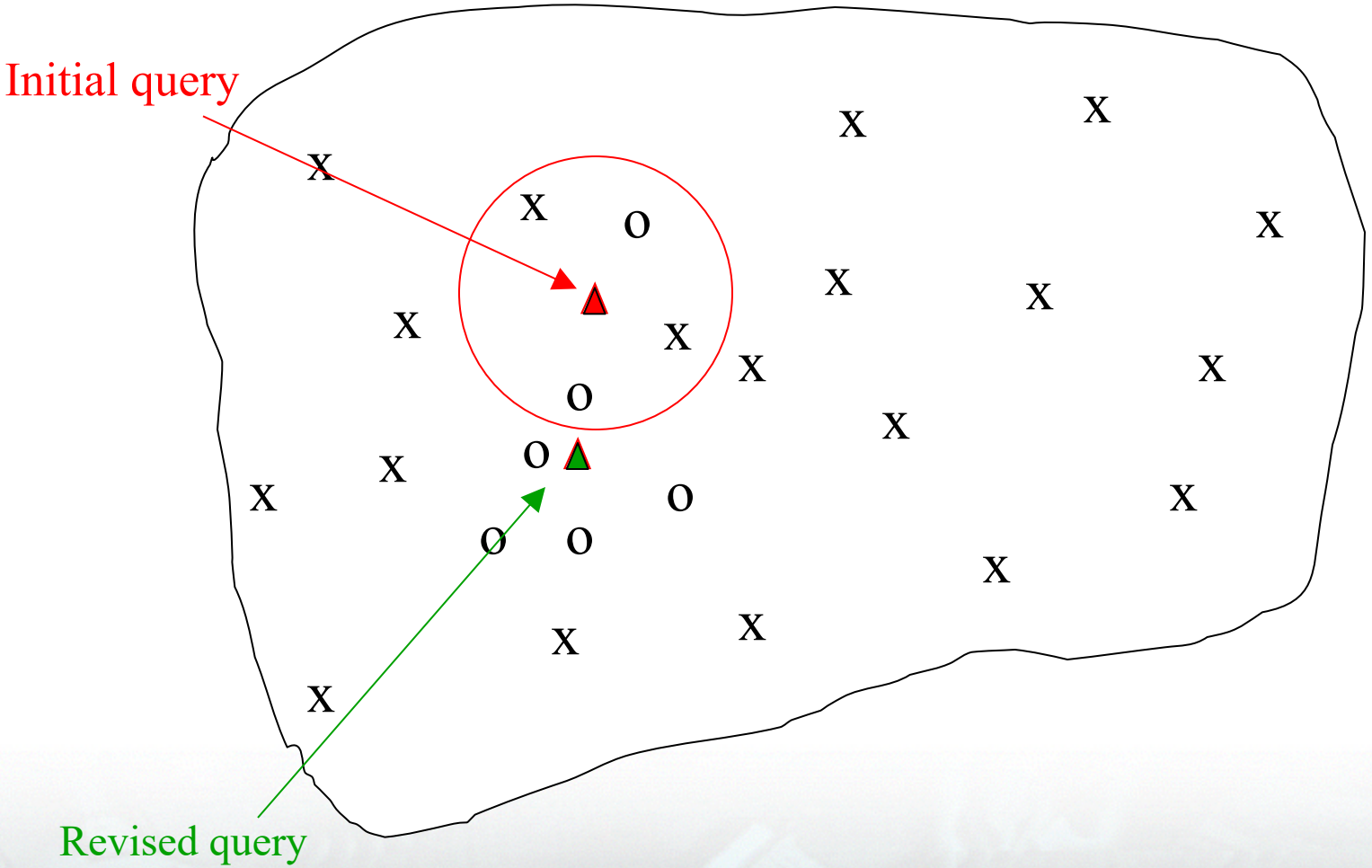


北京大学

Updating Queries

- Let's assume that there is an optimal query
  - The goal of relevance feedback is to bring the user query closer to the optimal query
- How does relevance feedback actually work?
  - Use relevance information to update query
  - Use query to retrieve new set of documents
- What exactly do we "feed back"?
  - Boost weights of terms from relevant documents
  - Add terms from relevant documents to the query
  - Note that this is hidden from the user

Picture of Relevance Feedback



X：non-relevant documents；O：relevant documents

北京大学

Rocchio Algorithm

- Used in practice:

$$\vec{q}_m = \alpha \vec{q}_0 + \beta \frac{1}{|D_r|} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$$

q_m = modified query vector;
q_0 = original query vector;
α,β,γ: weights (hand-chosen or set empirically);
D_r = set of known relevant doc vectors;
D_{nr} = set of known irrelevant doc vectors

- New query
  - Moves toward relevant documents
  - Away from irrelevant documents

Rocchio, J. Relevance feedback in information retrieval. The SMART Retrieval System: Experiments in Automatic Document Processing. G. Salton, ed. (1971). Prentice-Hall, Englewood Cliffs, NJ, 313–323.

Positive/Negative Relevance Feedback

- Positive RF (正反馈)
- Negative RF (负反馈)

- 正反馈价值往往大于负反馈
  - 例：可以通过设置 $\beta = 0.75, \gamma = 0.25$来给正反馈更大的权重
  - 很多系统甚至只允许正反馈，即$\gamma=0$

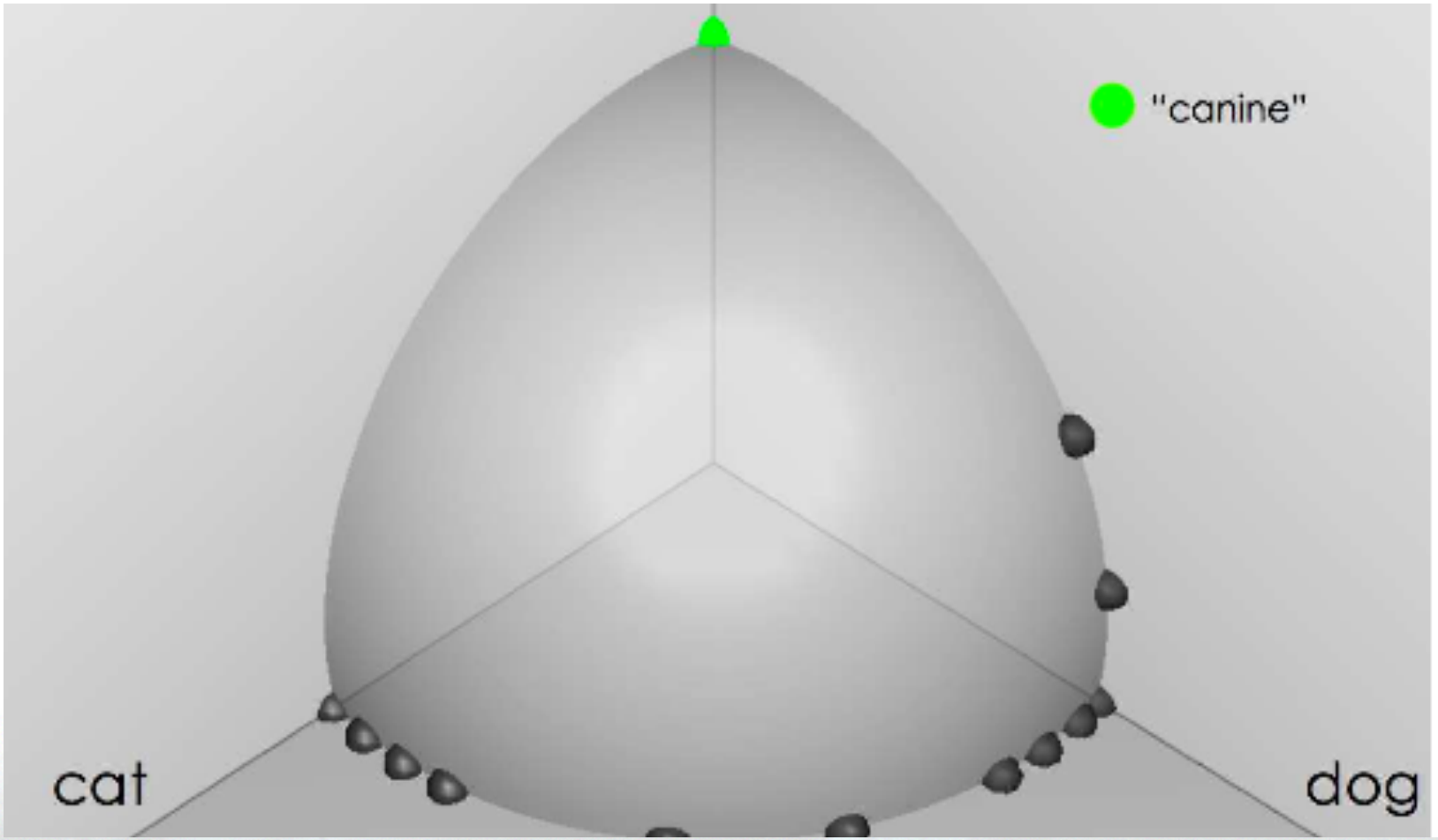
Rocchio in Pictures

$$\text{query vector} = \alpha \cdot \text{original query vector} + \beta \cdot \text{positive feedback vector} - \gamma \cdot \text{negative feedback vector}$$

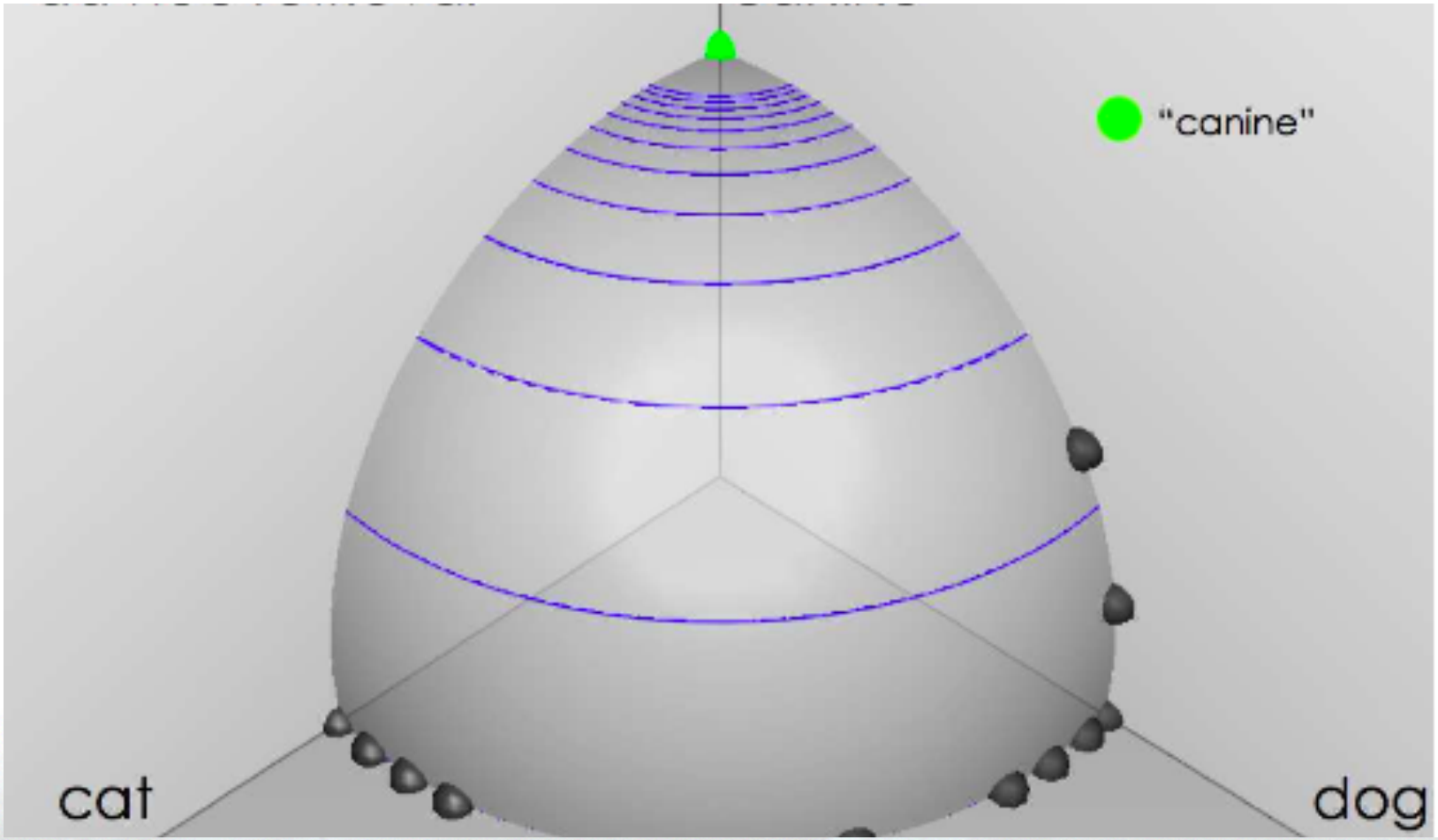
Typically, $\gamma < \beta$

Original query	0	4	0	8	0	0	$\alpha = 1.0$	0	4	0	8	0	0
Positive Feedback	2	4	8	0	0	2	$\beta = 0.5$	1	2	4	0	0	1	(+)
Negative feedback	8	0	4	4	0	16	$\gamma = 0.25$	2	0	1	1	0	4	(-)
New query								-1	6	3	7	0	-3

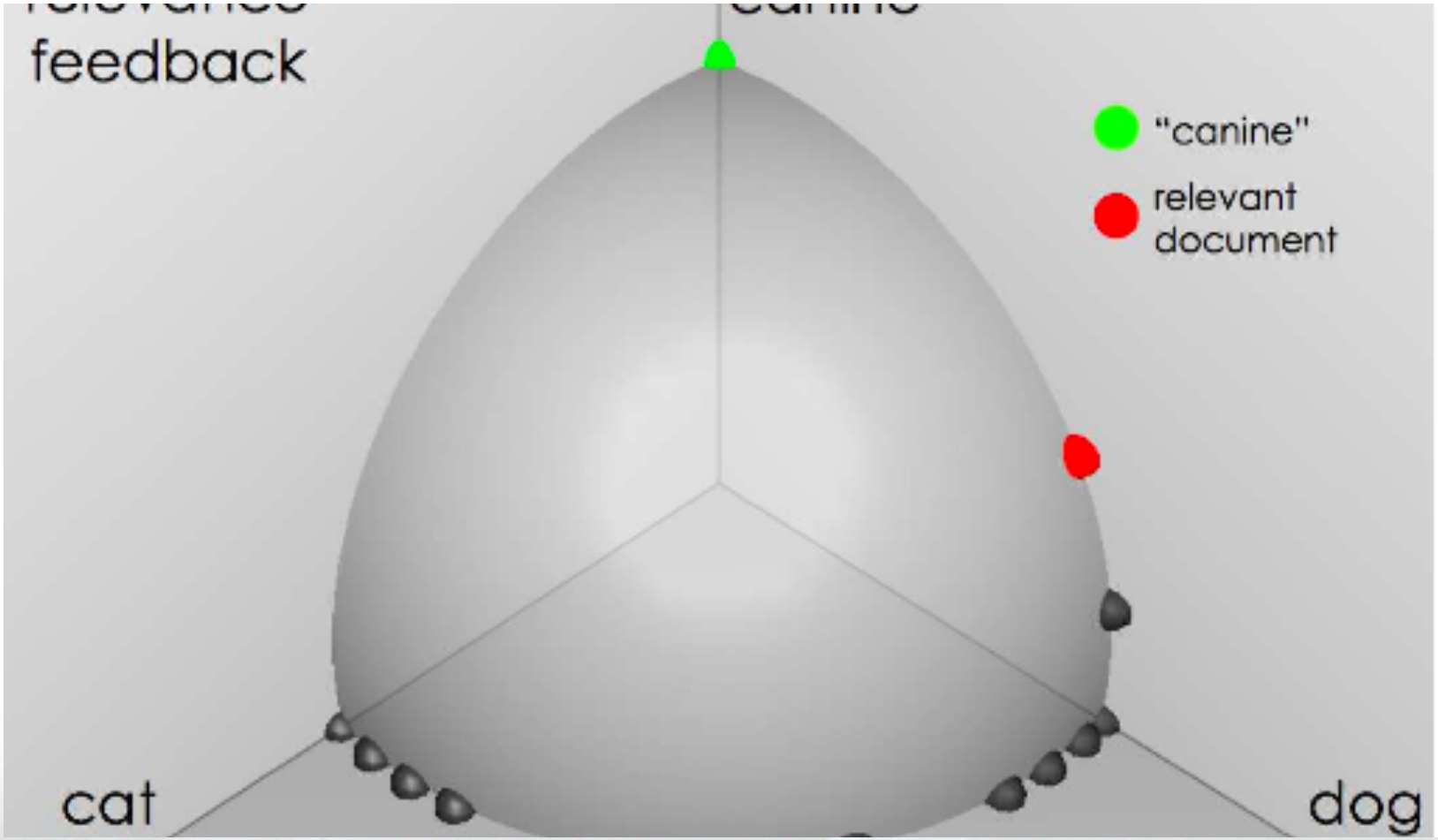
向量空间的例子: 查询 “canine”



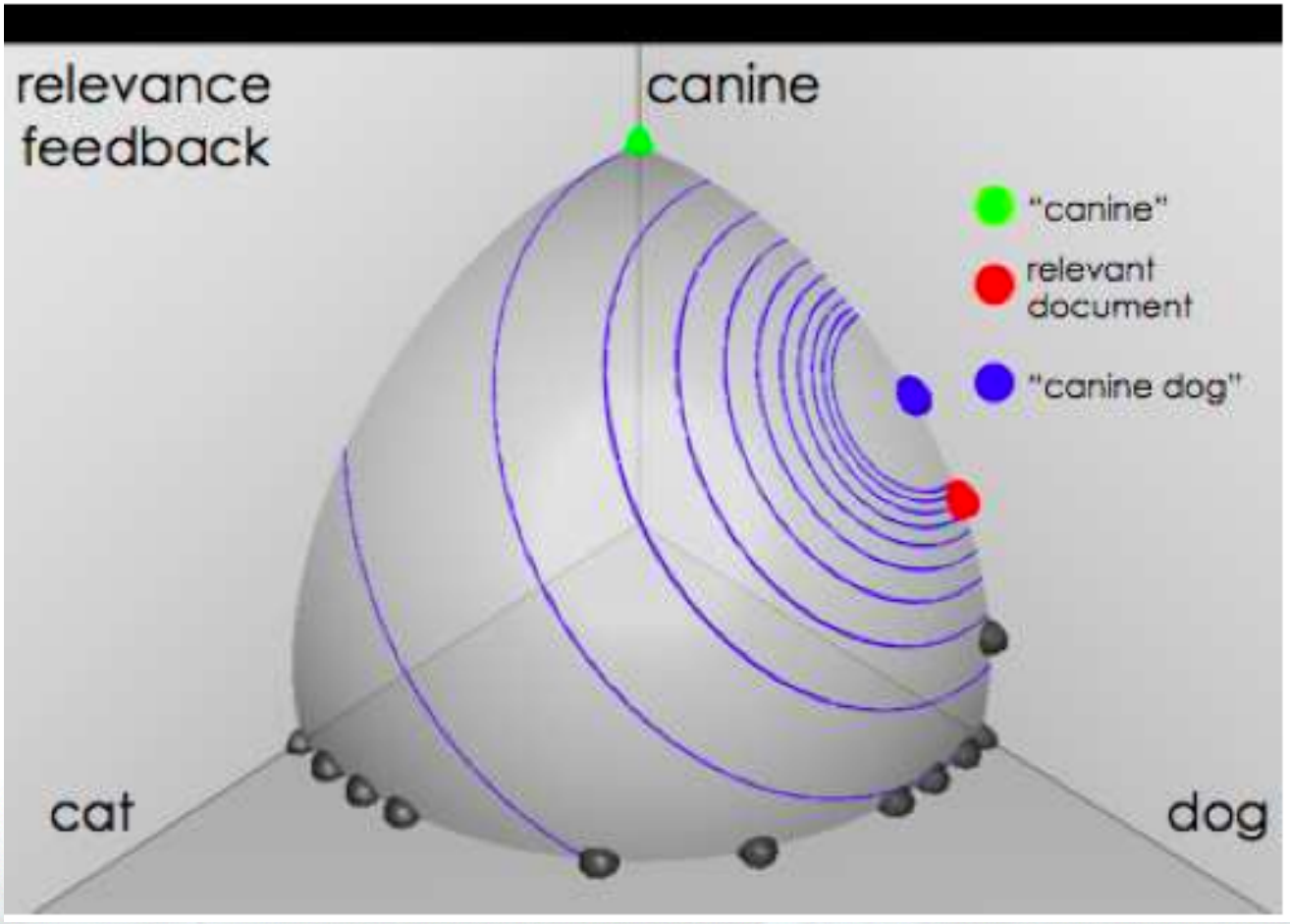
文档和查询 “canine” 的相似度



用户反馈: 选择相关文档



相关反馈后的检索结果

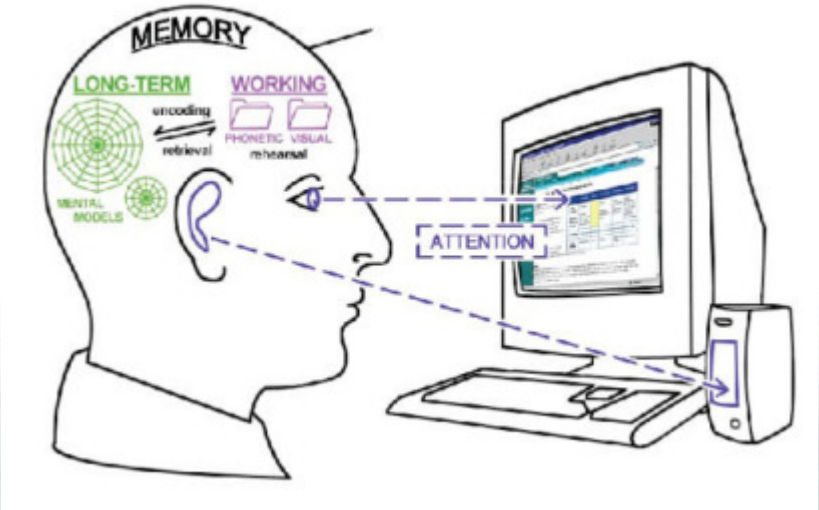

relevance
feedback
canine
"canine"
relevant
document
"canine dog"
cat
dog

Challenges to ERF

(1) Cognitive load
(2) Additional effort
(3) Failure to handle complex or multi-topic documents
(4) Nature of relevance judgments
(5) Reliance on initial result ranking
(6) Need for large quantities of feedback
(7) Assessment of documents individually by users
(8) The cost of explicit relevance feedback systems

(1) Cognitive Load

- Searchers may lack the cognitive capacities to manage the additional requirements of marking relevant documents effectively while trying to complete their search task.



北京大学

(2) Additional Effort

- The feedback mechanism is not implemented as part of the routine search activity; searchers may forget to use the feature or find it too onerous.

(3) Failure to handle complex or multi-topic documents

- The effectiveness of relevance feedback systems can suffer if the corpus has many multi-topic or partially relevant documents.
- RF systems treat documents as single entities with an inherent notion of relevance and non-relevance encompassing the whole entity, not the constituent parts.

(4) The nature of relevance judgments

- Relevance assessments are usually binary in nature, and seldom is partial relevance considered.
- Sliding scale

(5) Reliance on initial result ranking

- The searcher is only able to judge the relevance of the retrieved documents.
- If result precision is low, then users have limited opportunity to provide feedback.

(6) The need for large quantities of feedback

- To operate effectively, RF algorithms need large quantities of relevance information; but searchers appear reluctant to provide RF.

(7) Assessment of documents individually by users

- Incremental feedback requires searchers to assess documents individually; they are asked about the relevance of a document before being shown the next document.

(8) The cost of ERF

- Cost in time and effort expended by the searcher. Reading and rating a large number of documents is a costly activity.
- If this benefit cannot be guaranteed, feedback approaches based on passive observational evidence may be more appropriate.

Implicit Feedback

- 隐性相关反馈
- Users are often reluctant to provide relevance judgments
  - Some searches are precision-oriented
  - They're lazy!
- Can we gather feedback without requiring the user to do anything?
- Idea: gather feedback from observed user behavior

Observable Behavior

Minimum Scope

Behavior Category	Segment	Object	Class
Examine	View Listen	Select	
Retain	Print	Bookmark Save Purchase Delete	Subscribe
Reference	Copy / paste Quote	Forward Reply Link Cite	
Annotate	Mark up	Rate Publish	Organize

Four types of behavioral evidence

- Examine behaviors: where a searcher studies a documents;
- Retain behaviors: where a searcher saves a document for later use;
- Reference behaviors: involve users linking all or part of a document to another document.
- Annotate behaviors: where the searcher intentially adds personal value to an information object.

		Signal Type		
		Attention	Action	Content-based
Search stage	Before search			• Previous queries
	On search result pages	• Time on result list before first click • Total time on result lists • Query Reformulation Interval time	• Issue query: • Click-through: • Click order • Click position • Number of clicks • The way in which the user exited the page	• Query content features
	On content pages	• Display (dwell) time • First dwell time • Eye movement	• Scroll • Mouse movement & clicks • Number of visits	
	Further use of content pages		• Print; Bookmark; Save; Delete; Add to favorites; Email; Cite; Rate; Edit; etc.	

Dwell time as an indicator

- Positive relationship:
  - The longer the dwell time, the more likely the page is interested to the user.

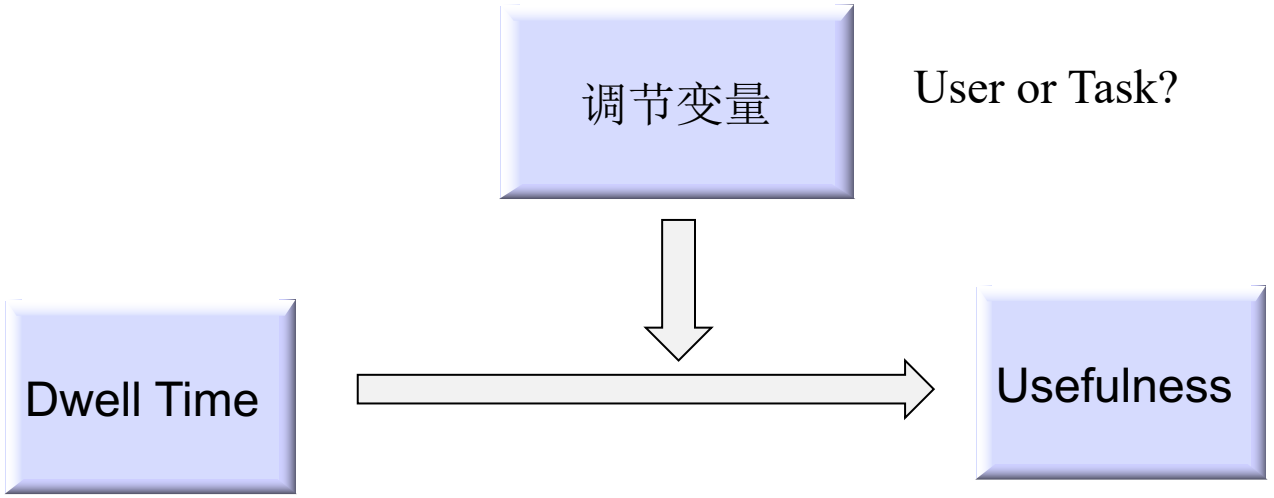
Morita and Shinoda (1994)

- No relationship between display time and usefulness judgments.

Kelly and Belkin (2004)

Dwell time as an indicator

- Interactions between dwell time, usefulness and the factors of user and task respectively.

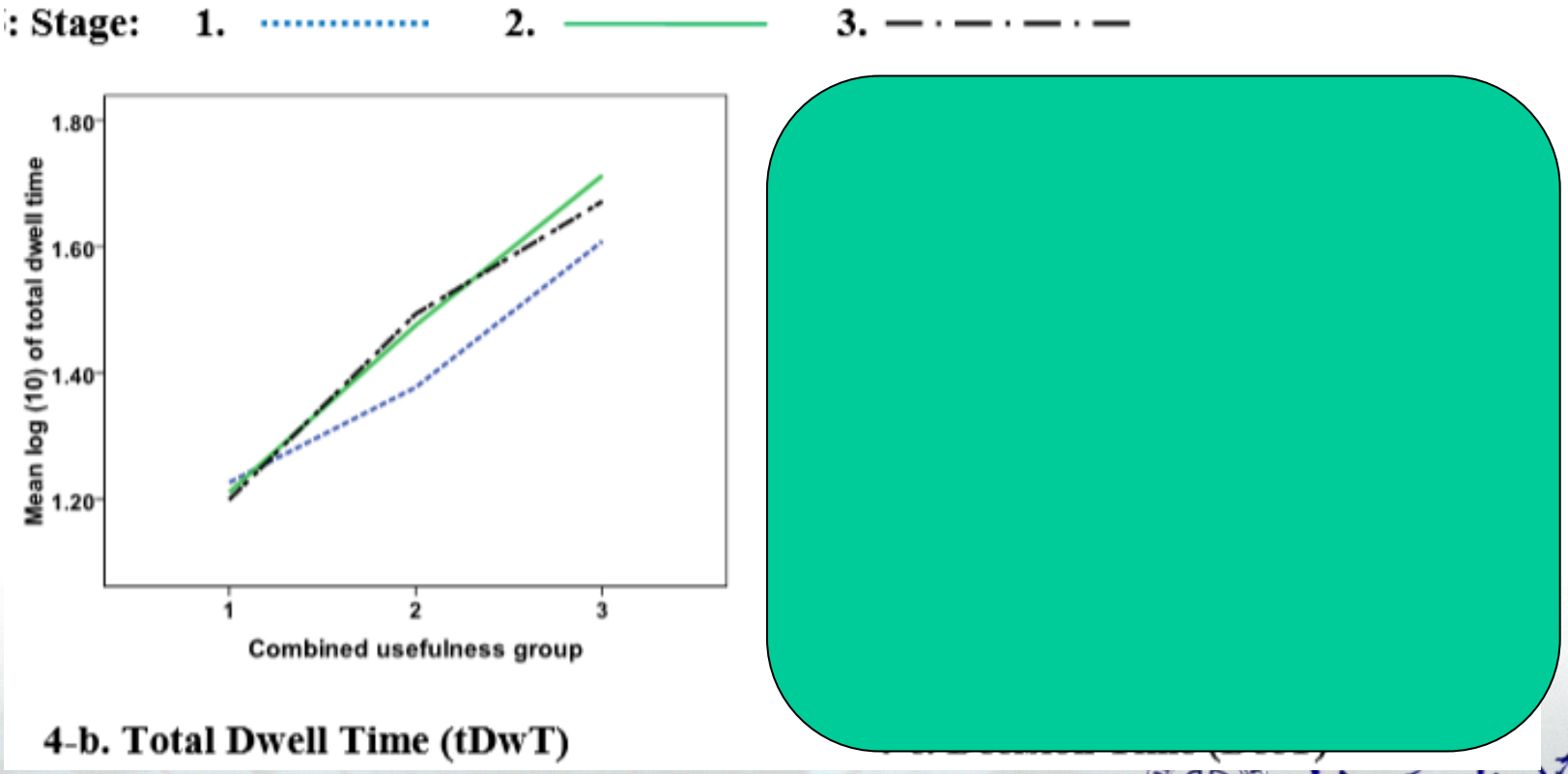


Threshold? Median?

White and Kelly (2006)

Dwell time as an indicator

- Interactions between dwell time, usefulness and the factors of search stage respectively.



4-b. Total Dwell Time (tDwT)

Koenemann and Belkin's Work

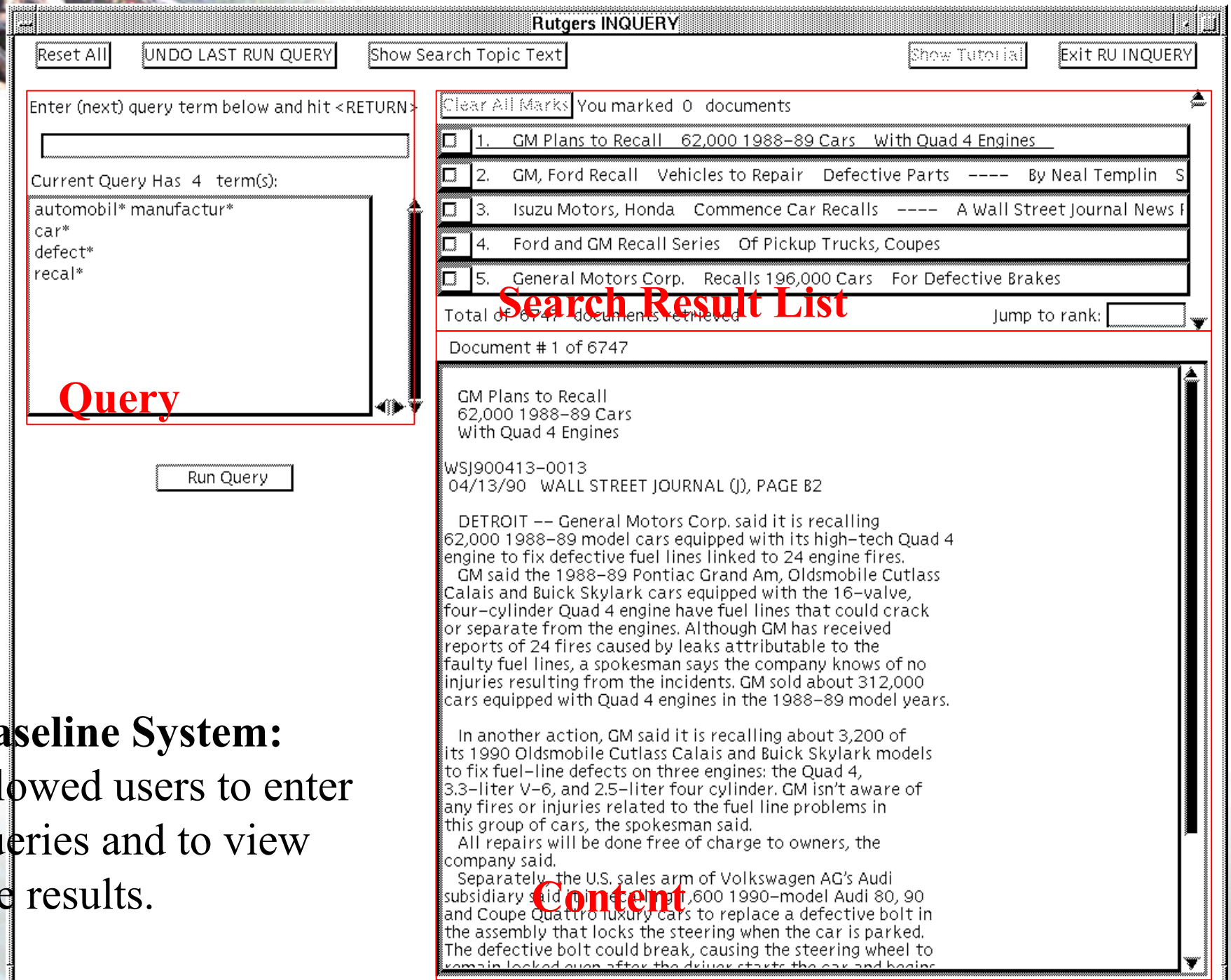
- Well-known study on relevance feedback in information retrieval
- Questions asked:
  - *How much knowledge and control should a user have in order to best interact with components such as relevance feedback that are central to the user task?*

Jürgen Koenemann and Nicholas J. Belkin. (1996) A Case For Interaction: A Study of Interactive Information Retrieval Behavior and Effectiveness. *Proceedings of SIGCHI 1996 Conference on Human Factors in Computing Systems (CHI 1996)*. Retrieved from http://www.sigchi.org/chi96/proceedings/papers/Koenemann/jk1_txt.htm

What is the best interface?

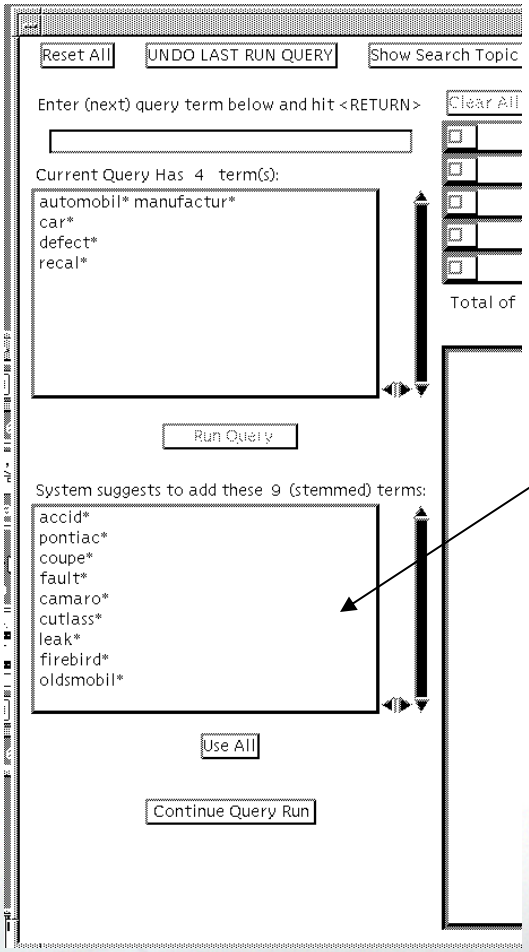
- Opaque (black box)
  - User doesn't get to see the relevance feedback process
- Transparent
  - User shown relevance feedback terms, but isn't allowed to modify query
- Penetrable
  - User shown relevance feedback terms and is allowed to modify the query

Which do you think worked best?



Baseline System:
allowed users to enter queries and to view the results.

Penetrable Interface



Users get to select which terms they want to add

北京大学

What is the best interface?

- Opaque (black box)
  - User doesn't get to see the relevance feedback process
- Transparent
  - User shown relevance feedback terms, but isn't allowed to modify query
- Penetrable
  - User shown relevance feedback terms and is allowed to modify the query

Which do you think worked best?

Research Questions

- Does relevance feedback improve results?
- Is user control over relevance feedback helpful?
- How do different levels of user control effect results?

Study Details

- Subjects started with a tutorial
  - 64 novice searchers (43 female, 21 male)
- Goal is to keep modifying the query until they have developed one that gets high precision
- INQUERY system used
- TREC collection (Wall Street Journal subset)
- Two search topics:
  - Automobile Recalls
  - Tobacco Advertising and the Young
- Relevance judgments from TREC and experimenter

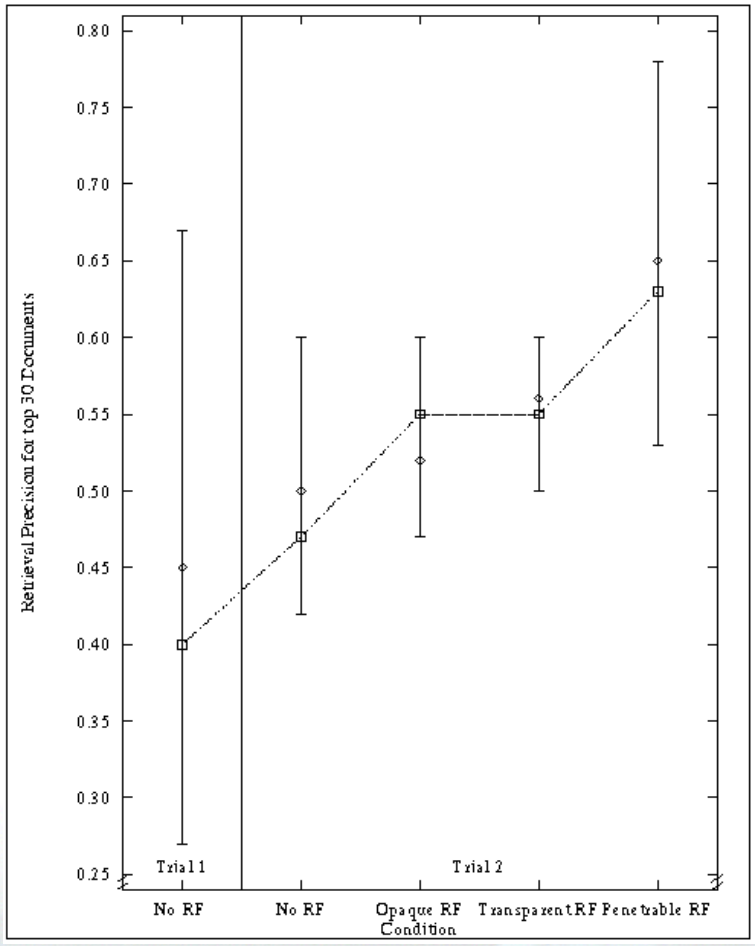
Sample Topic

Topic: Tobacco company advertising and the young
Description: A document will provide information on what is a widely held opinion that the tobacco industry aims its advertising at the young.
Narrative: A relevant document must report on tobacco company advertising and its relation to young people. A relevant document can address either side of the question: (1) Do tobacco companies consciously target the young, or (2) As the tobacco industry argues, is this an erroneous public perception. The "young" may be identified as youth, children, adolescents, teenagers, high school students, and college students.

Procedure

- Baseline (Trial 1)
  - Subjects get tutorial on relevance feedback
- Experimental condition (Trial 2)
  - Shown one of four modes: no relevance feedback, opaque, transparent, penetrable
- Evaluation metric used: precision at 30 documents

Precision Results

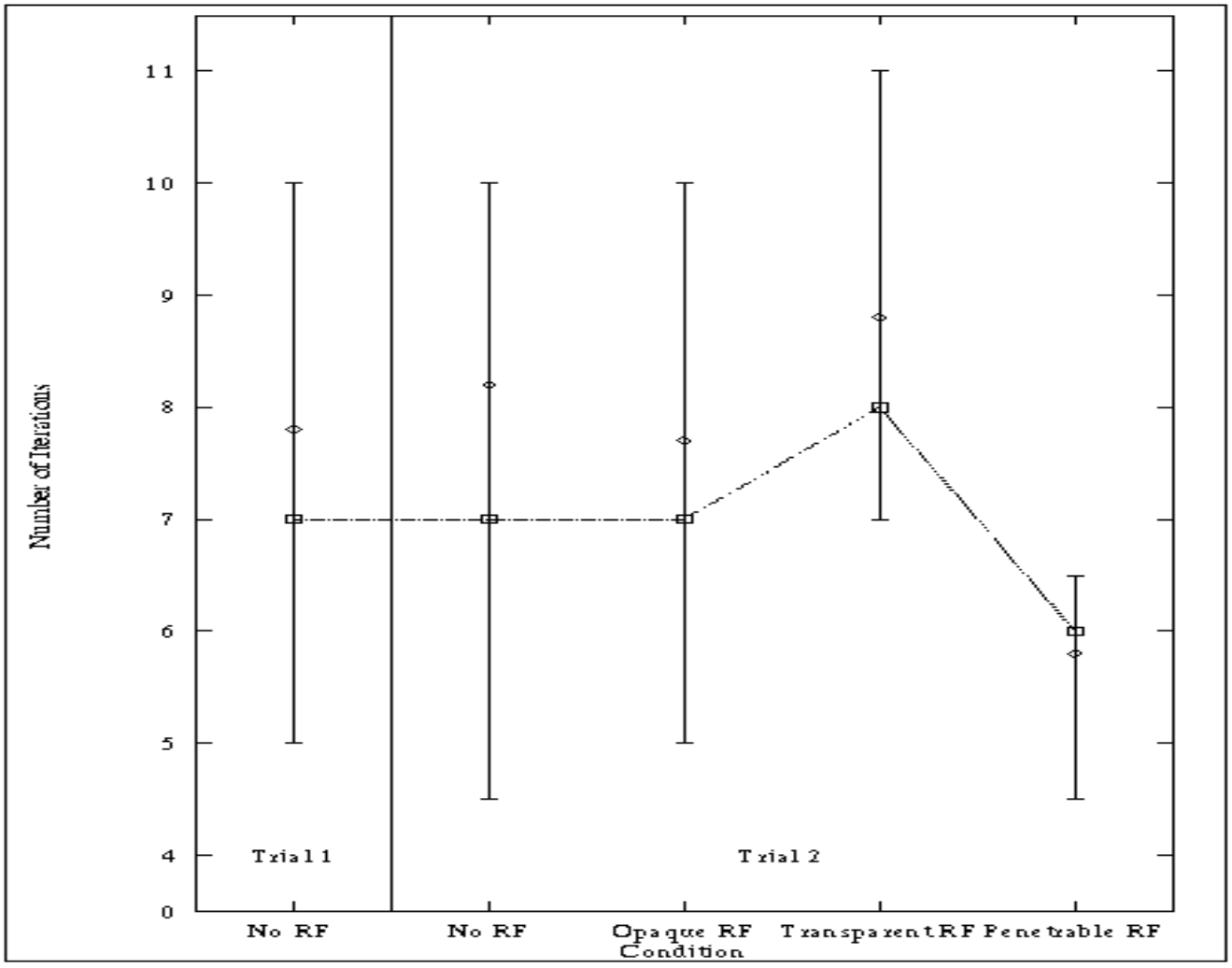


北京大学

Relevance feedback works!

- Subjects using the relevance feedback interfaces performed 17-34% better
- Subjects in the penetrable condition performed 15% better than those in opaque and transparent conditions

Number of Iterations



北京大学

Behavior Results

- Search times approximately equal
- Precision increased in first few iterations
- Penetrable interface required fewer iterations to arrive at final query
- Queries with relevance feedback are much longer
  - But fewer terms with the penetrable interface
  - users were more selective about which terms to add

Pseudo Relevance Feedback

- Also called “blind relevance feedback”
- 伪相关反馈，也称之为盲式相关反馈
- Motivation: it’s difficult to elicit relevance judgments from users
- Idea: take top n documents, and simply assume that they are relevant
- Perform relevance feedback as before
- If the initial hit list is reasonable, system should pick up good query terms

PRF Experiment

- Retrieval engine: Indri
- Test collection: TREC, topics 301-450
- Procedure:
  - Used topic description as query to generate initial hit list
  - Selected top 20 terms from top 20 hits using tf.idf
  - Added these terms to the original query

PRF Example

Number: 303
Title: Hubble Telescope Achievements

Description:
Identify positive accomplishments of the Hubble telescope since it was launched in 1991.

Narrative:
Documents are relevant that show the Hubble telescope has produced new data, better quality data than previously available, data that has increased human knowledge of the universe, or data that has led to disproving previously existing theories or hypotheses. Documents limited to the shortcomings of the telescope would be irrelevant. Details of repairs or modifications to the telescope without reference to positive achievements would not be relevant.

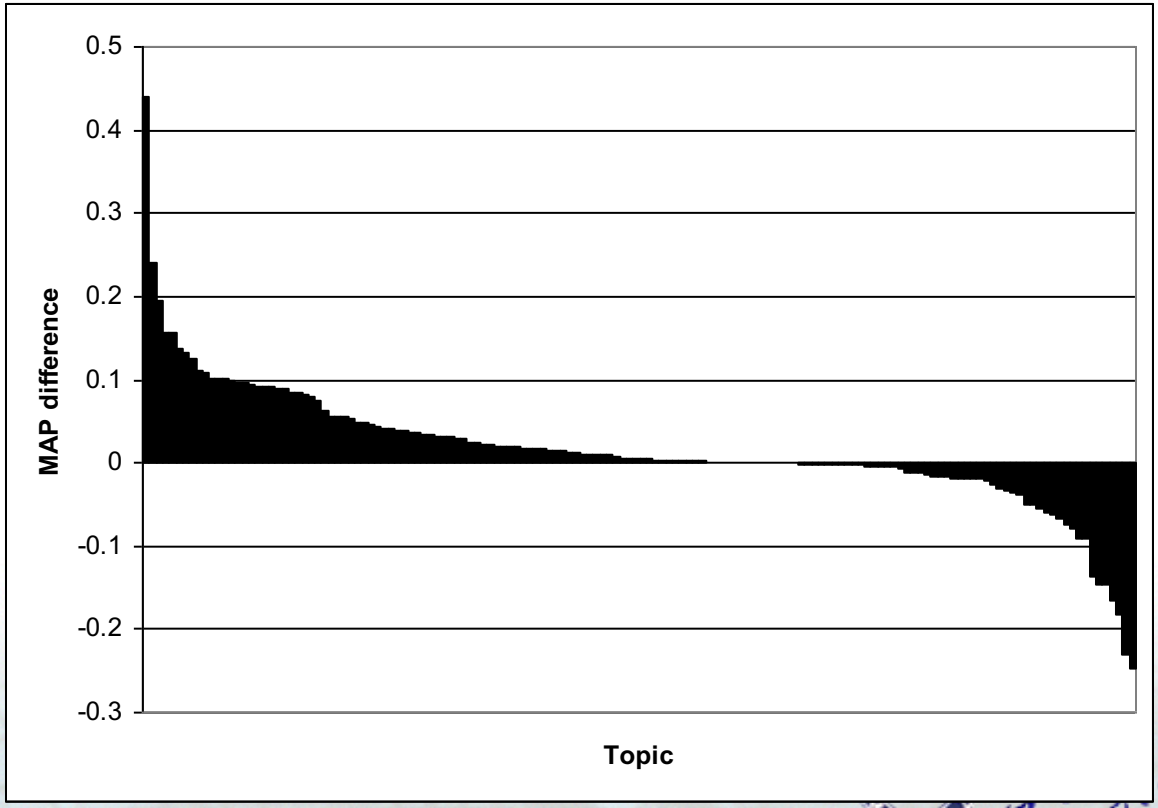
Terms added

telescope	1041.33984032195
hubble	573.896477205696
space	354.090789112131
nasa	346.475671454331
ultraviolet	242.588034029191
shuttle	230.448255669841
mirror	184.794966339329
telescopes	155.290920607708
earth	148.865466409231
discovery	146.718067628756
orbit	142.597040178043
flaw	141.832019493907
scientists	132.384677410089
launch	116.322861618261
stars	116.205713485691
universe	114.705686405825
mirrors	113.677943638299
light	113.59717006967
optica zl	106.198288687586
species	103.555123536418

Results

	MAP	R-Precision
No feedback	0.1591	0.2022
With feedback	0.1806 (+13.5%)	0.2222 (+9.9%)

Pseudo relevance feedback does not always help!



Kelly's study

- What kind of query suggestions do users prefer?
  - suggested query? suggested term?
  - system suggestion? user suggestion?

Kelly, D., Gyllstrom, K., & Bailey, E. W. A comparison of term and query suggestion features for interactive searching. Proc. SIGIR 2009, ACM Press, (2009), 371-378.

Independent Variables

- Suggestions type:
  - Query suggestion system
  - Term suggestion system
  - Within-subject design
- Source of suggestions:
  - SGS=system-generated suggestions
  - UGS=user-generated suggestions
  - Between subject design

Query Suggestion System

北京大学

Term Suggestion System

北京大学

Dependent Variables

- Users' usage of Suggestions
- Performance:
  - number of documents saved
  - session-based normalized discounted cumulated gain (sDCG)
- Users' perceptions & preferences
  - effectiveness, satisfaction and preference in Exit Questionnaire

Procedure

- 55 subjects （33 females and 22 males）
- 20 topics
- Each subject completed four topics, half with a term suggestion system and half with a query suggestion system.
- Before search: the consent form, demographic questionnaire and search experience questionnaire.
- After search: exit questionnaire

Results: suggestion type

- Subjects used more query suggestions than term suggestions and saved more documents with these suggestions.
- Subjects preferred the query suggestion system and rated it higher, e.g., its ability to help them think of new approaches to searching.

Results: suggestion source

- Subjects who received user-generated suggestions saved more documents found through suggestions; the most were saved for query suggestions.
- The performance measure (snDCG) showed that the best performance was achieved by those who received user-generated query suggestions.

Results: interaction effect

- The performance measure (snDCG) showed that there appeared to be an interaction with source of suggestions and suggestion type:
  - those who received user-generated suggestions did better with query suggestions,
  - those with system-generated terms did better with term suggestions.

本节内容参考的教材章节

- Ruthven, I., Kelly, D. Chapter 9. Interactive techniques.

https://shimo.im/forms/Vnq7Attdyx07y7Y4/fill
《交互式信息检索》第11讲内容反馈