Lost in publications?

How to find your way in 50 million scientific documents

Based on: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster*, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013. And later publications including: Jaakko Peltonen, Jonathan Strahl, and Patrik Floreen. Negative Relevance Feedback for Exploratory Search with Visual Interactive Intent Modeling. In Proceedings of IUI 2017


CONTRAST
BRIGHT
NAV MODE
CAT
INS
AHRS
AM
IMU
DISPLAY
ALT NUM
SYM ELEM
DATA LINK
JAN STROBE
NON ATTK
VEL VECTOR
TID MODE
GND STAB
A/C STAB
RANGE
100
50

From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

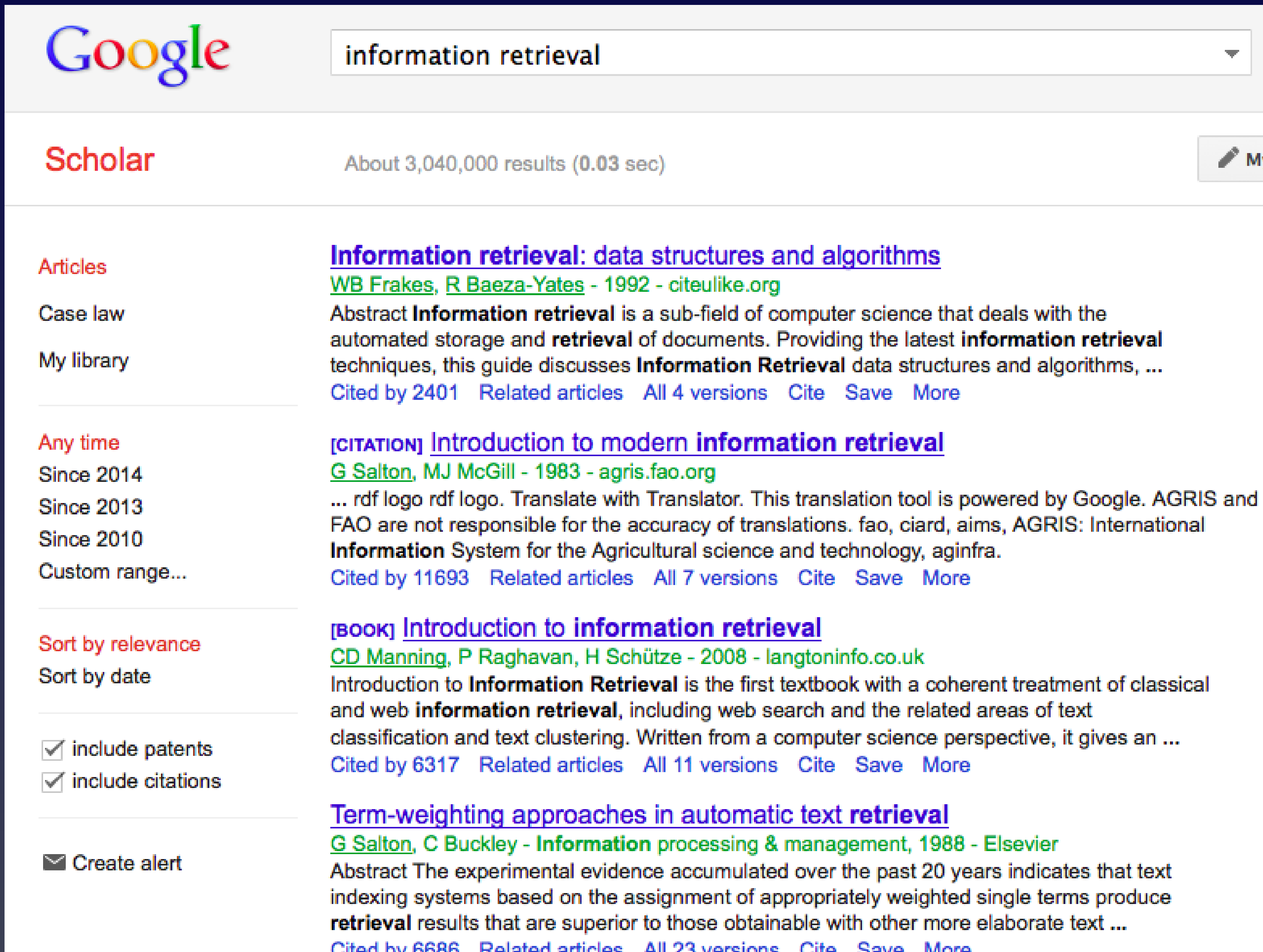
Information Seeking: Overview

Researchers must **navigate big data**. Current scientific knowledge includes 50 million published articles. How can a system **help a researcher find relevant documents** in her field?

We introduce **IntentRadar**, an interactive search user interface and search engine that **anticipates user's search intents** by estimating them from user's **interaction** with the interface. The estimated intents are **visualized** on a radial layout that organizes potential intents as directions in the information space.

From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

Typical interfaces for scientific search



From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

A main problem: hard to formulate queries precisely, information needs evolve.

Search engines can mistake what you are looking for.

You may not know what precisely you are looking for, or may not be able to express it as a search phrase.

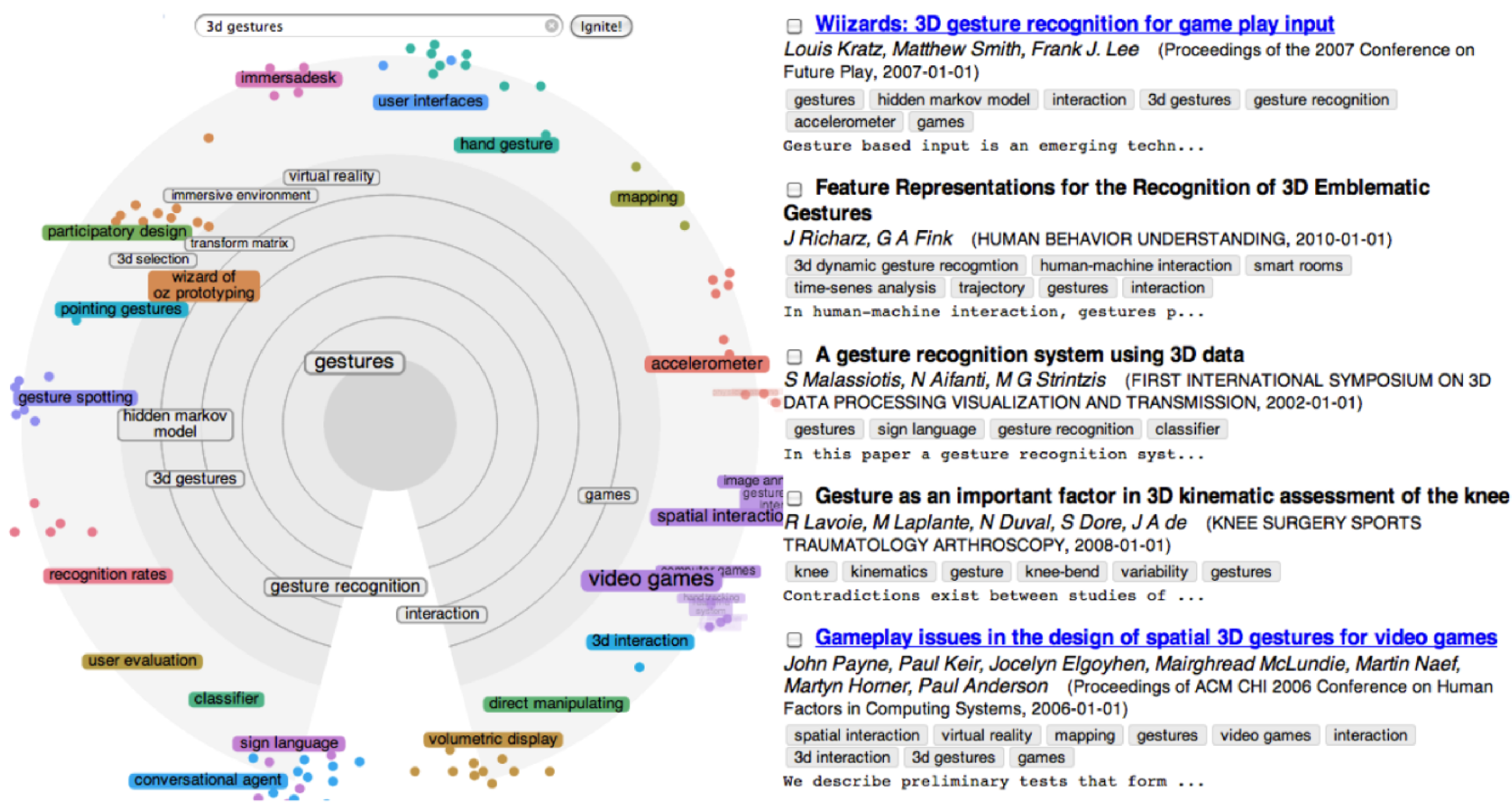
There is a disconnect between what the computer thinks you need, and what you actually need.

Traditional interfaces only allow you to try a search phrase, and try again if you don't like the results.

----> “guessing game”: what phrase (if any) will give the results I need

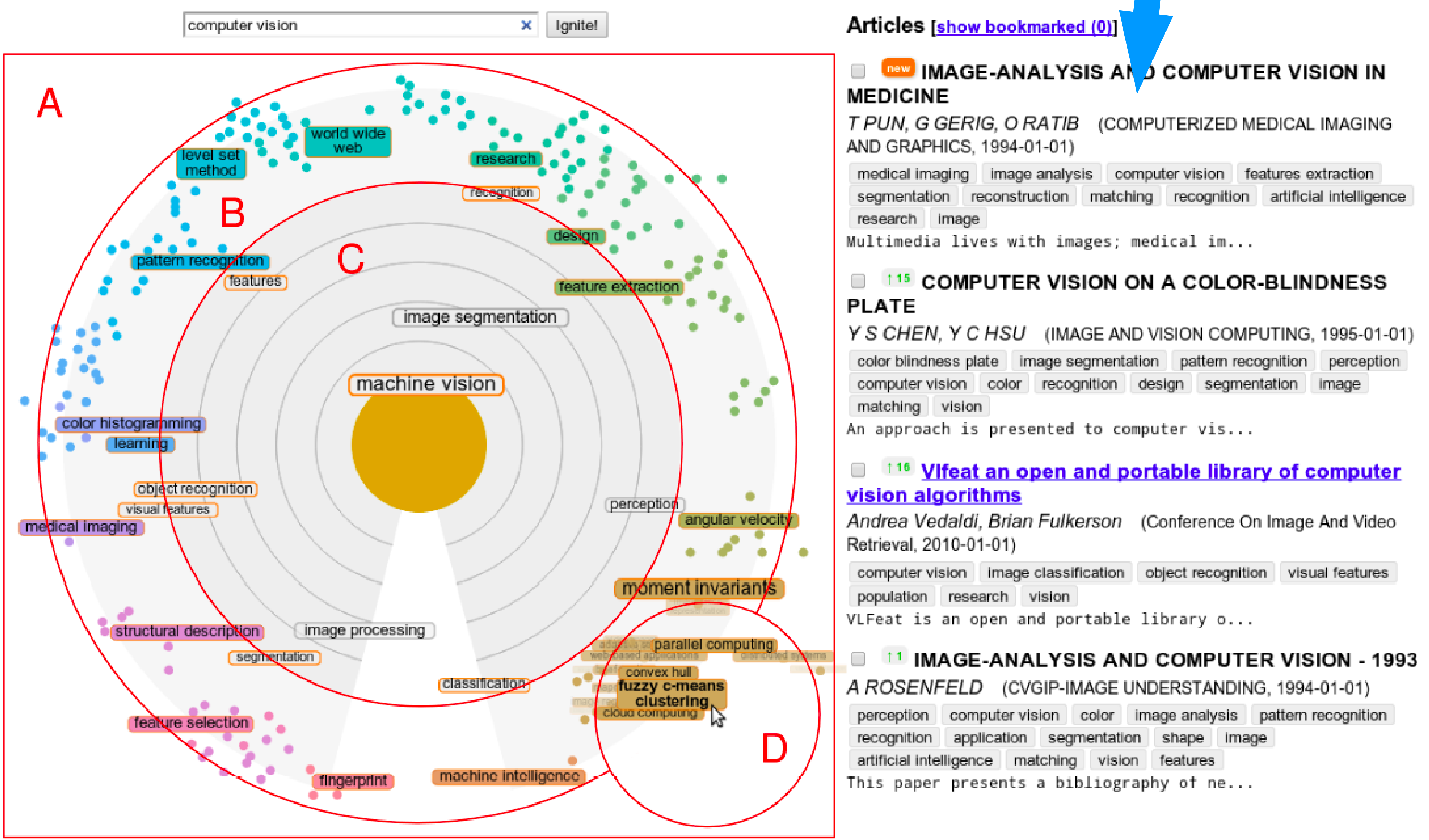
From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

Our approach: Radar layout



From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

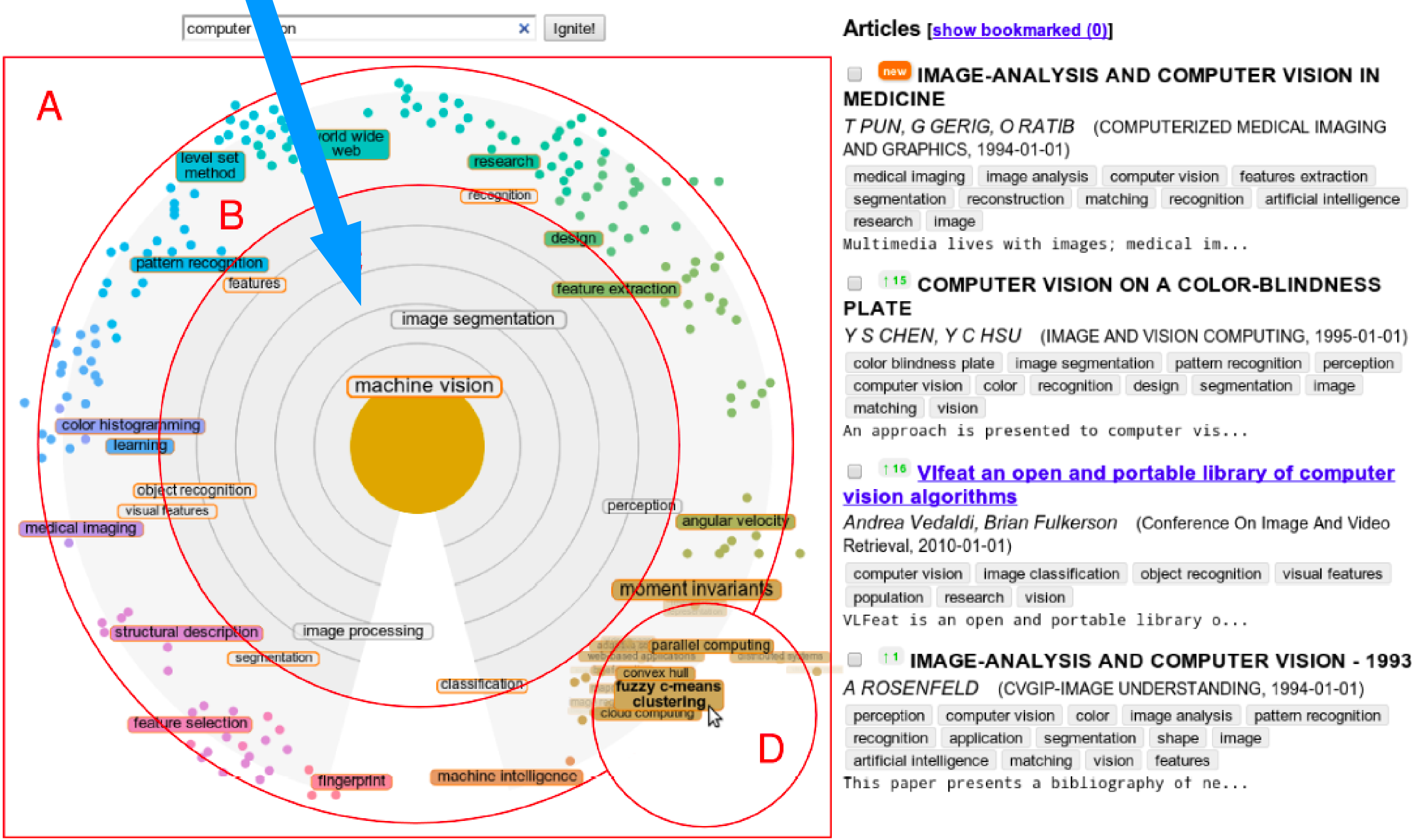
Article list



Current intent estimation for which results are retrieved.

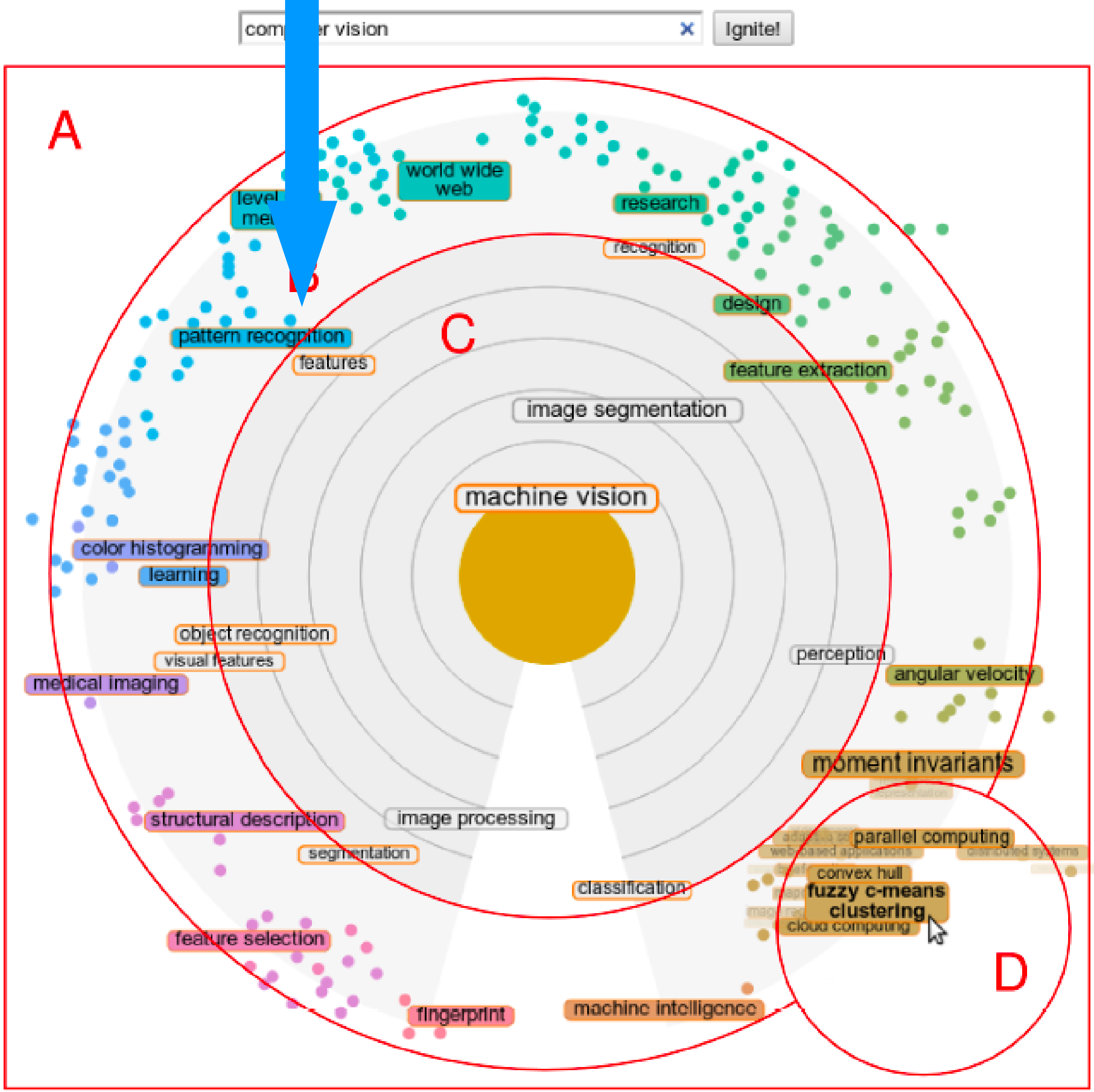
Angular distance = similarity of intents, radius = relevance

From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.



Predicted intents (help users to find directions on the radar to move away from their currently estimated intent)

From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.



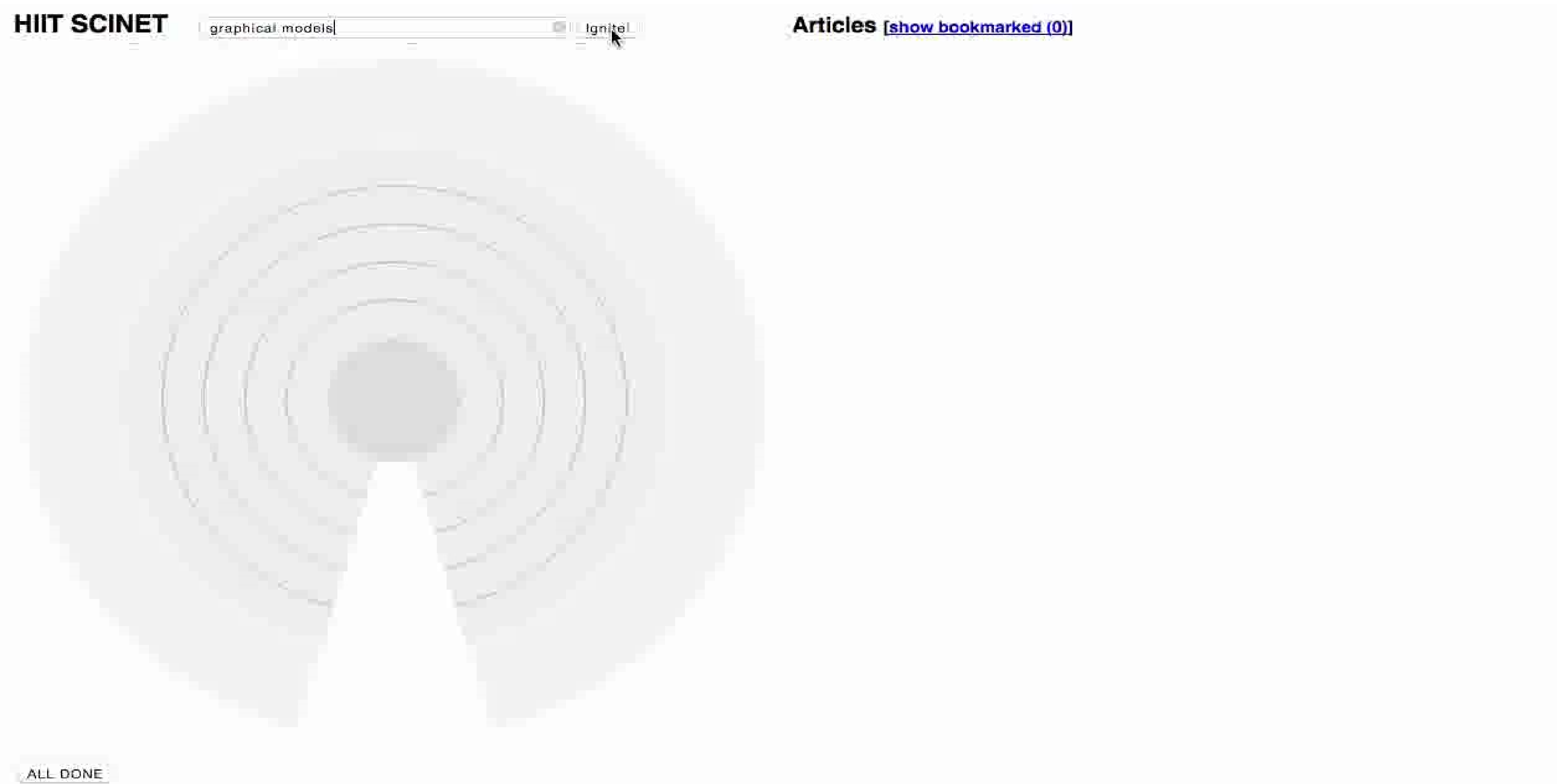
Articles [show bookmarked (0)]

new **IMAGE-ANALYSIS AND COMPUTER VISION IN MEDICINE**
T PUN, G GERIG, O RATIB (COMPUTERIZED MEDICAL IMAGING AND GRAPHICS, 1994-01-01)
medical imaging | image analysis | computer vision | features extraction | segmentation | reconstruction | matching | recognition | artificial intelligence | research | image
Multimedia lives with images; medical im...

↑15 **COMPUTER VISION ON A COLOR-BLINDNESS PLATE**
Y S CHEN, Y C HSU (IMAGE AND VISION COMPUTING, 1995-01-01)
color blindness plate | image segmentation | pattern recognition | perception | computer vision | color | recognition | design | segmentation | image | matching | vision
An approach is presented to computer vis...

↑16 **Vlfeat an open and portable library of computer vision algorithms**
Andrea Vedaldi, Brian Fulkerson (Conference On Image And Video Retrieval, 2010-01-01)
computer vision | image classification | object recognition | visual features | population | research | vision
VLFeat is an open and portable library o...

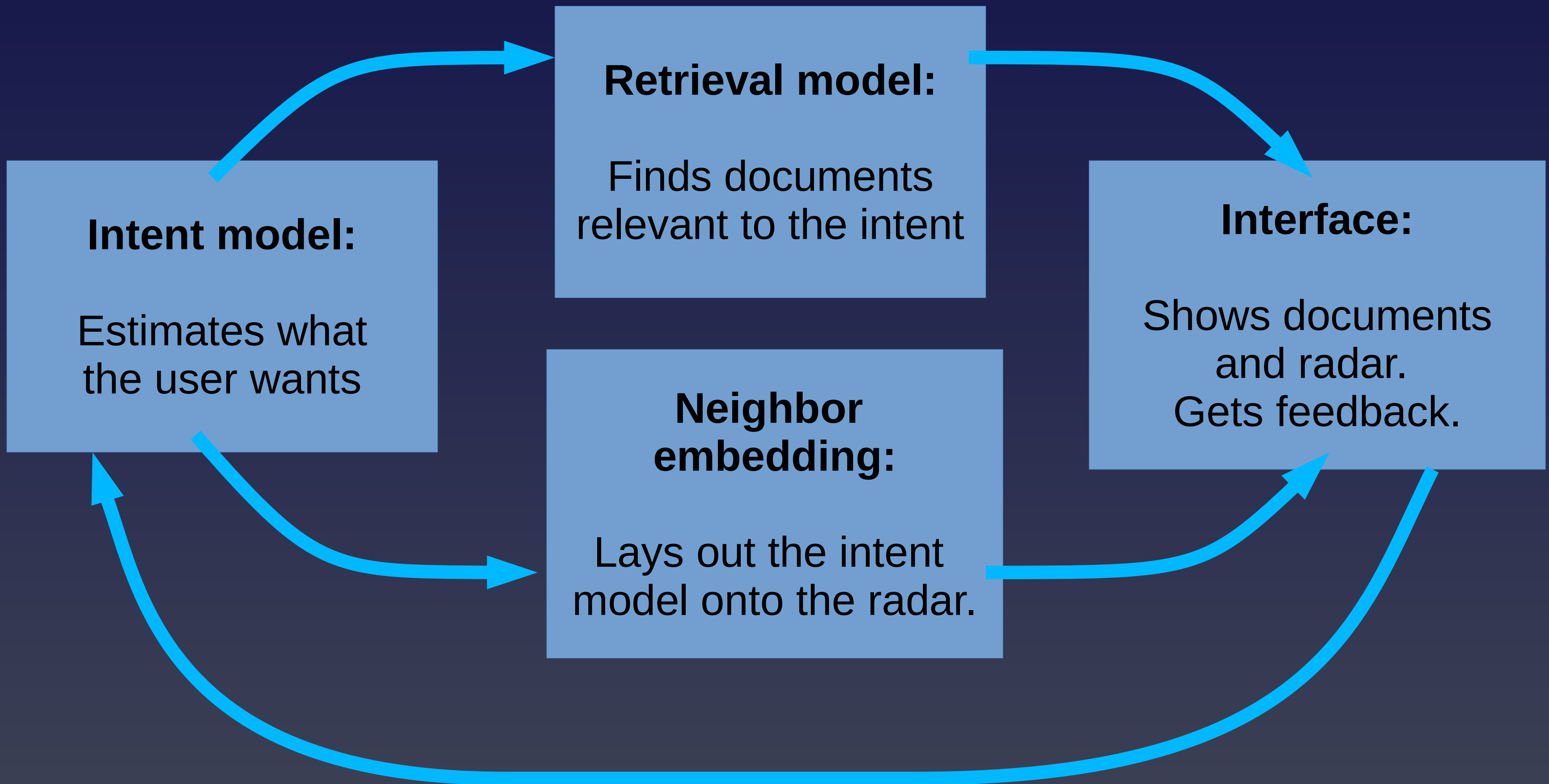
↑1 **IMAGE-ANALYSIS AND COMPUTER VISION - 1993**
A ROSENFELD (CVGIP-IMAGE UNDERSTANDING, 1994-01-01)
perception | computer vision | color | image analysis | pattern recognition | recognition | application | segmentation | shape | image | artificial intelligence | matching | vision | features
This paper presents a bibliography of ne...


HIIT SCINET
graphical models
Ignite!
Articles [show bookmarked (0)]
ALL DONE

From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

Machine learning for the Intent Radar

- Learning of user's search intents during interactive search
- Based on a **retrieval model** and **intent model**; layout based on **neighbor embedding**



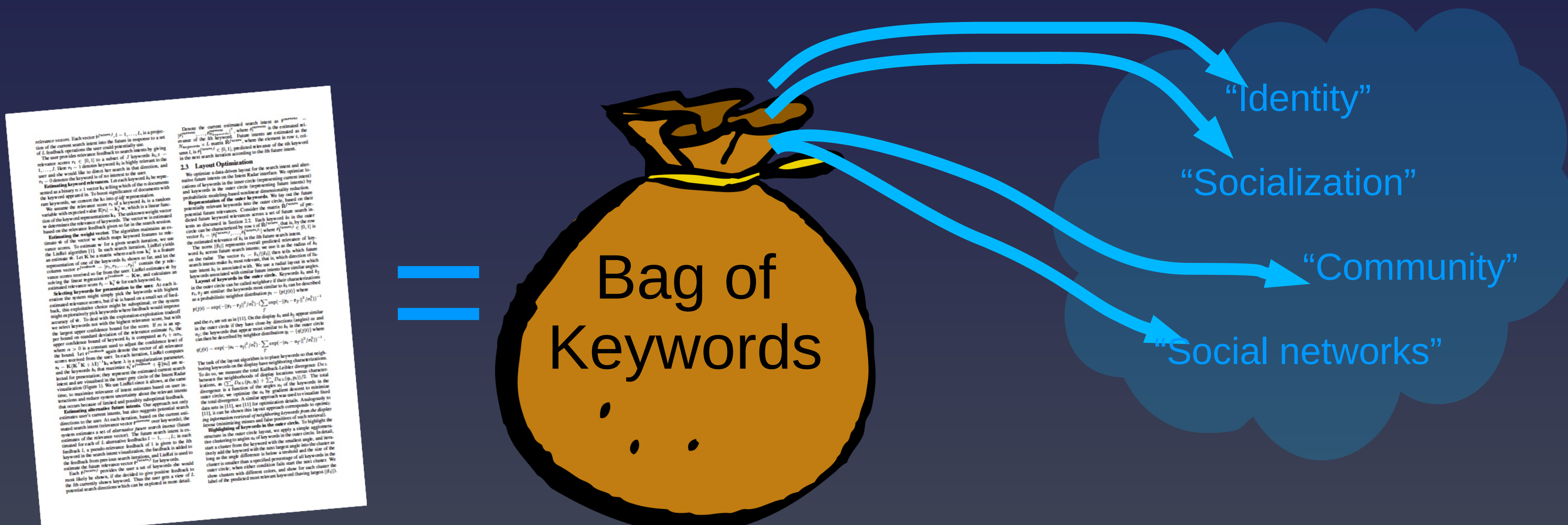
From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

Retrieval Model

- Estimates probability of relevant documents based on estimates of the intent model

- We use the language modeling approach of information retrieval

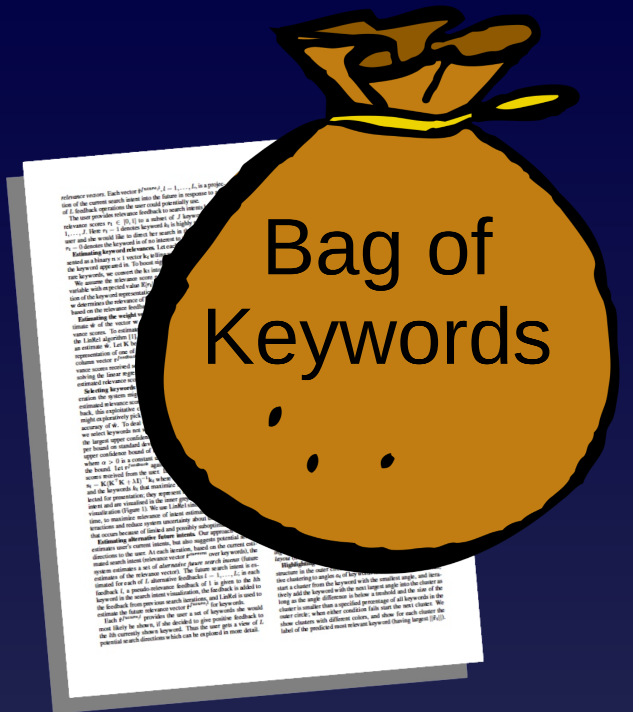
- Unigram language model, Bayesian Dirichlet Smoothing



Document retrieval

Estimated user intent model used for retrieval:

Top positive keyword predictions treated as a sample of an **ideal desired document**



Documents ranked by **likelihood** to produce desired document

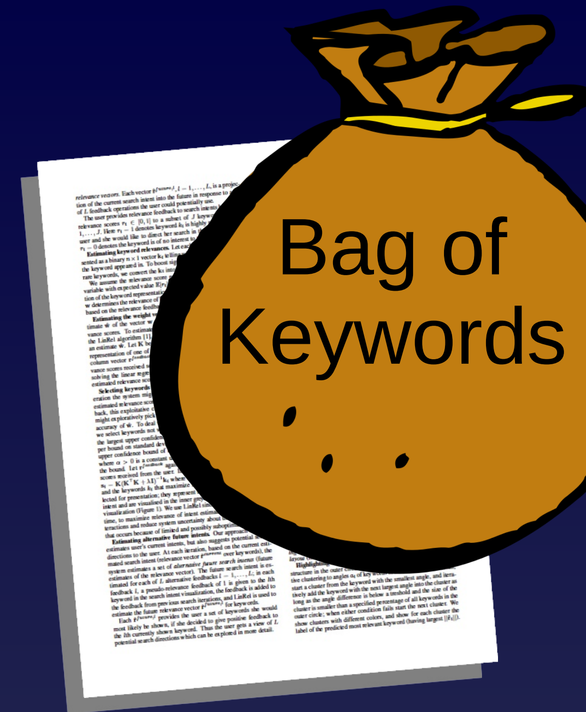
$$score(d) = \log P(\mathbf{v}_P | M_d)$$

$$= \sum_{i \in K_P} v_i^P \log P(v_i^P | M_d)$$

Likelihood: is the document likely to produce keywords of ideal desired content?

Dirichlet sampling exposes user to more novel documents.

Document retrieval



Estimated user intent model used for retrieval:

Top positive keyword predictions treated as a sample of an **ideal desired document**

Top negative keyword predictions treated as a sample of an **ideal unwanted document**

Documents ranked by **likelihood ratio** to produce desired document rather than unwanted document

$$score(d) = \log \frac{P(\mathbf{v}_P | M_d)}{P(\mathbf{v}_N | M_d)}$$

$$= \sum_{i \in K_P} v_i^P \log P(v_i^P | M_d) - \sum_{i \in K_N} v_i^N \log P(v_i^N | M_d)$$

Likelihood ratio: is the document more likely to produce keywords of ideal desired content or of unwanted content?

Dirichlet sampling exposes user to more novel documents.

From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

Intent Model

- Estimates **current search intent** and **alternative future intents** that could occur in response to user feedback

- We use the **LinRel algorithm**. Yields estimate of keyword weights in each iteration, based on interaction history.

- Observations = relevance scores given by user to keywords. Assumption: expected relevance = linear function of what documents the keyword appears in.

$$\mathbf{r}^{feedback} = [r_1, r_2, \ldots, r_p]^\top$$

Feedback scores in [0,1] given so far to a subset of keywords

$$\mathbf{r}^{feedback} = \mathbf{K}\mathbf{w}$$

Model feedback: regression based on what documents they appeared in (matrix **K**)

$$\hat{r}_i = \mathbf{k}_i^\top \hat{\mathbf{w}}$$

Use model to estimate relevance of the rest of the keywords

Keyword relevance prediction

- Relevance of keywords is estimated by Bayesian inference (Bayesian linear regression incl. a propagation step)
- Balance of exploration & exploitation: keywords with **highest Upper Confidence Bounds** are shown in inner circle; ones with **lowest Lower Confidence Bounds** in the outer ring

$$E[r_i] = \mathbf{k}_i^\top \mathbf{w} = \mathbf{s}_i^\top \mathbf{r}^{feedback} +$$

Expected relevance of a keyword: linear regression of what documents it appears in. Solution: linear function of keyword feedbacks so far.

$$\mathbf{s}_i = \mathbf{K}(\mathbf{K}^\top \mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{k}_i$$

Linear relevance predictor for one keyword, based on its co-occurrence in the same documents as keywords that have received direct feedback.

$$\mathbf{s}_i^\top \mathbf{r}^{feedback} + \frac{\alpha}{2} ||s_i||$$

Upper Confidence Bound: highest relevance the keyword might reasonably be expected to get. Balances current best estimate and remaining uncertainty.

$$\mathbf{s}_i^\top \mathbf{r}^{feedback} - \frac{\alpha}{2} ||s_i||$$

Lower Confidence Bound: lowest relevance the keyword might reasonably be expected to get.

From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

Layout of Intents by Nonlinear Dimensionality Reduction

- Radial position of each keyword = current estimated relevance

- Angles are used to represent directions of future intent

- Each keyword is represented by its relevances in all future intents (high-dimensional representation):

$$\tilde{\mathbf{r}}_i = [\hat{r}_i^{future,l}, \ldots, \hat{r}_i^{future,L}]$$

Relevance of keyword i in future intent L

$$\bar{\mathbf{r}}_i = \tilde{\mathbf{r}}_i / ||\tilde{\mathbf{r}}_i||$$

Normalized vector, tells which future intents (or feedbacks) make keyword i most relevant

- Layout is **optimized for retrieval of keywords** with similar relevance in future intents, by **nonlinear dimensionality reduction.** We use a well-performing approach optimized for information retrieval, details in Wednesday's talk.

From: Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. Directing Exploratory Search with Interactive Intent Modeling. In Proceedings of CIKM 2013, ACM Conference on Information and Knowledge Management, pages 1759-1764. ACM, 2013.

User Experiment - Results