

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



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.

Typical interfaces for scientific search

The screenshot shows a search interface for 'information retrieval'. At the top left is the Google logo. To its right is a search bar containing the query 'information retrieval'. Below the search bar, the word 'Scholar' is displayed in red, indicating the search mode. To the right of 'Scholar' is a status message 'About 3,040,000 results (0.03 sec)'. Further right is a 'My' button with a pencil icon. On the far left, there is a sidebar with navigation links: 'Articles', 'Case law', 'My library', 'Any time' (with options for 'Since 2014', 'Since 2013', 'Since 2010', and 'Custom range...'), 'Sort by relevance' (selected), 'Sort by date', and checkboxes for 'include patents' and 'include citations'. At the bottom of the sidebar is a 'Create alert' button. The main content area lists several search results. The first result is a book titled 'Information retrieval: data structures and algorithms' by WB Frakes and R Baeza-Yates from 1992, available on citeulike.org. It includes a brief abstract, citation information ('Cited by 2401'), and links to related articles, versions, citation, save, and more. The second result is a book titled '[CITATION] Introduction to modern information retrieval' by G Salton and MJ McGill from 1983, available on agris.fao.org. It includes a note about the translation tool, citation information ('Cited by 11693'), and links to related articles, versions, citation, save, and more. The third result is a book titled '[BOOK] Introduction to information retrieval' by CD Manning, P Raghavan, and H Schütze from 2008, available on langtoninfo.co.uk. It includes a brief abstract, citation information ('Cited by 6317'), and links to related articles, versions, citation, save, and more. The fourth result is a book titled 'Term-weighting approaches in automatic text retrieval' by G Salton and C Buckley from 1988, available on Elsevier. It includes a brief abstract, citation information ('Cited by 6686'), and links to related articles, versions, citation, save, and more.

Google

information retrieval

Scholar

About 3,040,000 results (0.03 sec)

My

Articles

Case law

My library

Any time

Since 2014

Since 2013

Since 2010

Custom range...

Sort by relevance

Sort by date

include patents

include citations

Create alert

[\[CITATION\] Information retrieval: data structures and algorithms](#)
WB Frakes, R Baeza-Yates - 1992 - citeulike.org
Abstract **Information retrieval** is a sub-field of computer science that deals with the automated storage and **retrieval** of documents. Providing the latest **information retrieval** techniques, this guide discusses **Information Retrieval** data structures and algorithms, ...
Cited by 2401 Related articles All 4 versions Cite Save More

[\[BOOK\] Introduction to modern information retrieval](#)
G Salton, MJ McGill - 1983 - agris.fao.org
... rdf logo rdf logo. Translate with Translator. This translation tool is powered by Google. AGRIS and FAO are not responsible for the accuracy of translations. fao, ciard, aims, AGRIS: International **Information** System for the Agricultural science and technology, aginfra.
Cited by 11693 Related articles All 7 versions Cite Save More

[\[BOOK\] Introduction to information retrieval](#)
CD Manning, P Raghavan, H Schütze - 2008 - langtoninfo.co.uk
Introduction to **Information Retrieval** is the first textbook with a coherent treatment of classical and web **information retrieval**, including web search and the related areas of text classification and text clustering. Written from a computer science perspective, it gives an ...
Cited by 6317 Related articles All 11 versions Cite Save More

[Term-weighting approaches in automatic text retrieval](#)
G Salton, C Buckley - **Information processing & management**, 1988 - Elsevier
Abstract The experimental evidence accumulated over the past 20 years indicates that text indexing systems based on the assignment of appropriately weighted single terms produce **retrieval** results that are superior to those obtainable with other more elaborate text ...
Cited by 6686 Related articles All 23 versions Cite Save More

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

Search engines can mistake what you are looking for.

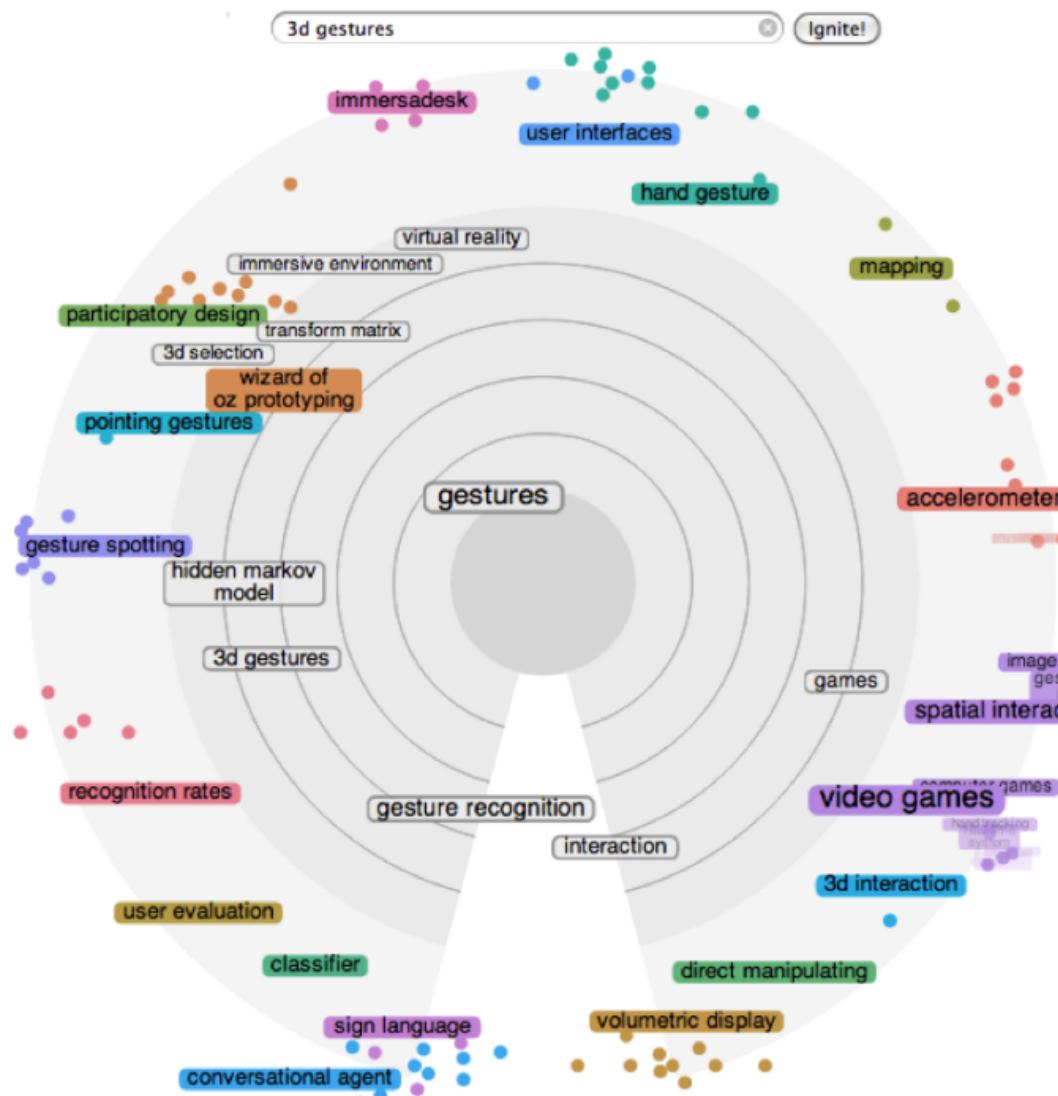
You may not know what precisely you are looking for, or may not be able to express it as a search phrase.

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.

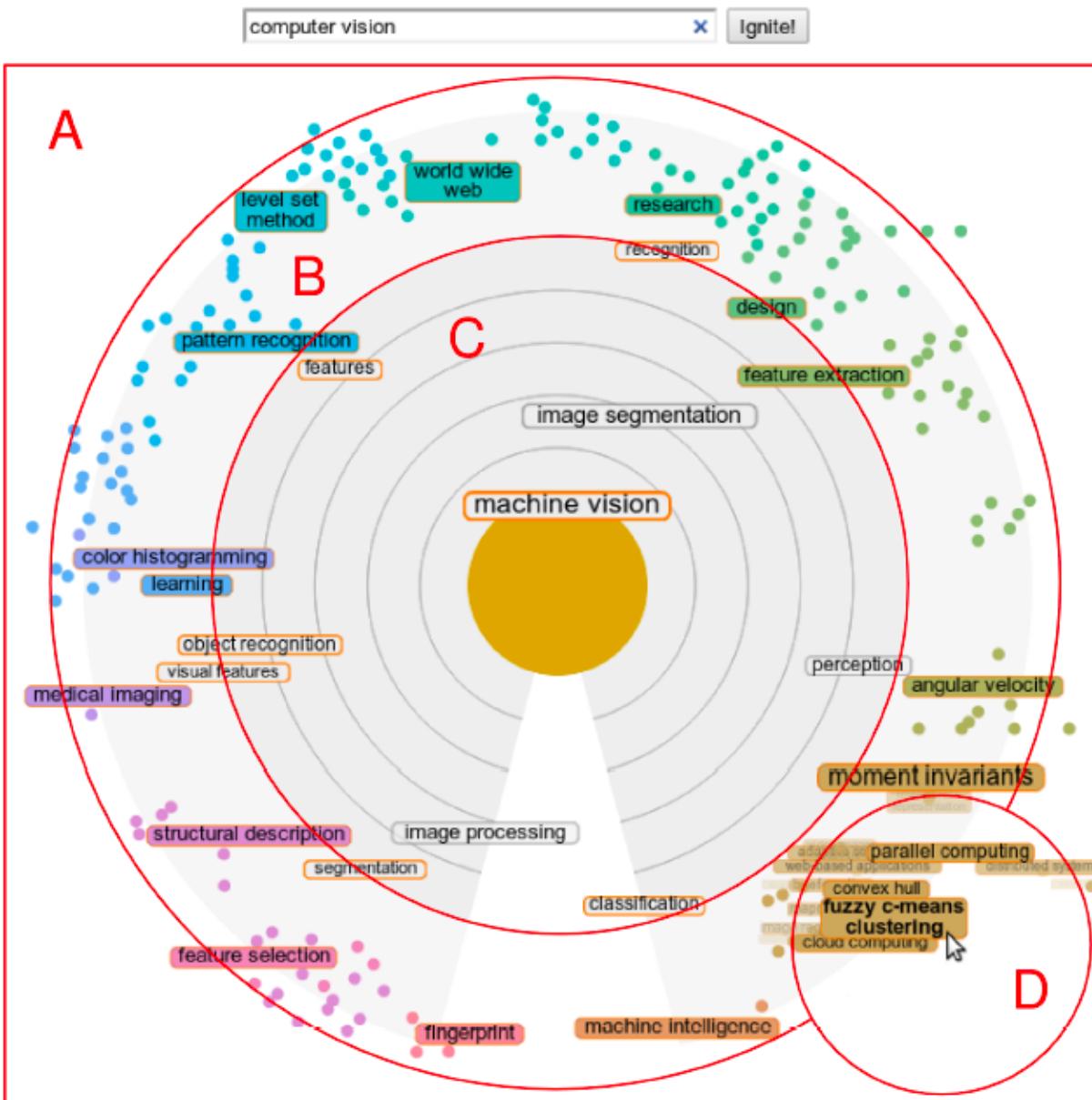
----> “guessing game”: what phrase (if any) will give the results I need

Our approach: Radar layout



- [**Wizards: 3D gesture recognition for game play input**](#)
Louis Kratz, Matthew Smith, Frank J. Lee (Proceedings of the 2007 Conference on Future Play, 2007-01-01)
gestures hidden markov model interaction 3d gestures gesture recognition accelerometer games
Gesture based input is an emerging techn...
 - Feature Representations for the Recognition of 3D Emblematic Gestures**
J Richarz, G A Fink (HUMAN BEHAVIOR UNDERSTANDING, 2010-01-01)
3d dynamic gesture recognition human-machine interaction smart rooms time-series analysis trajectory gestures interaction
In human-machine interaction, gestures p...
 - A gesture recognition system using 3D data**
S Malassiotis, N Aifanti, M G Strintzis (FIRST INTERNATIONAL SYMPOSIUM ON 3D DATA PROCESSING VISUALIZATION AND TRANSMISSION, 2002-01-01)
gestures sign language gesture recognition classifier
In this paper a gesture recognition syst...
 - Gesture as an important factor in 3D kinematic assessment of the knee**
R Lavoie, M Laplante, N Duval, S Dore, JA de (KNEE SURGERY SPORTS TRAUMATOLOGY ARTHROSCOPY, 2008-01-01)
knee kinematics gesture knee-bend variability gestures
Contradictions exist between studies of ...
 - Gameplay issues in the design of spatial 3D gestures for video games**
John Payne, Paul Keir, Jocelyn Elgoyhien, Mairghread McLundie, Martin Naef, Martyn Horner, Paul Anderson (Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems, 2006-01-01)
spatial interaction virtual reality mapping gestures video games interaction 3d interaction 3d gestures games
We describe preliminary tests that form ...

Article list



Articles [show bookmarked (0)]

1 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

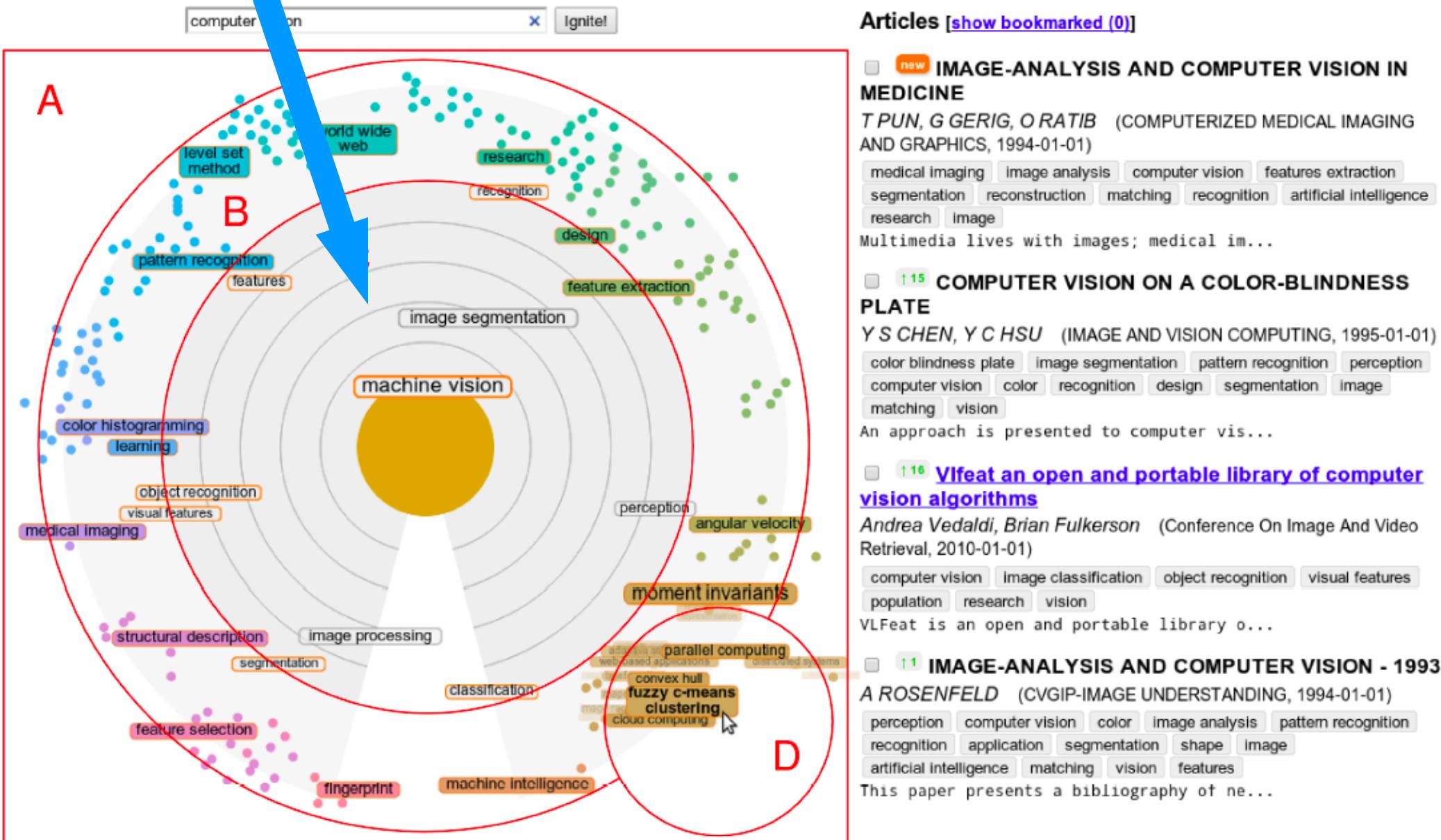
A ROSENFELD (CVGIP-IMAGE UNDERSTANDING, 1994-01-01)

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...

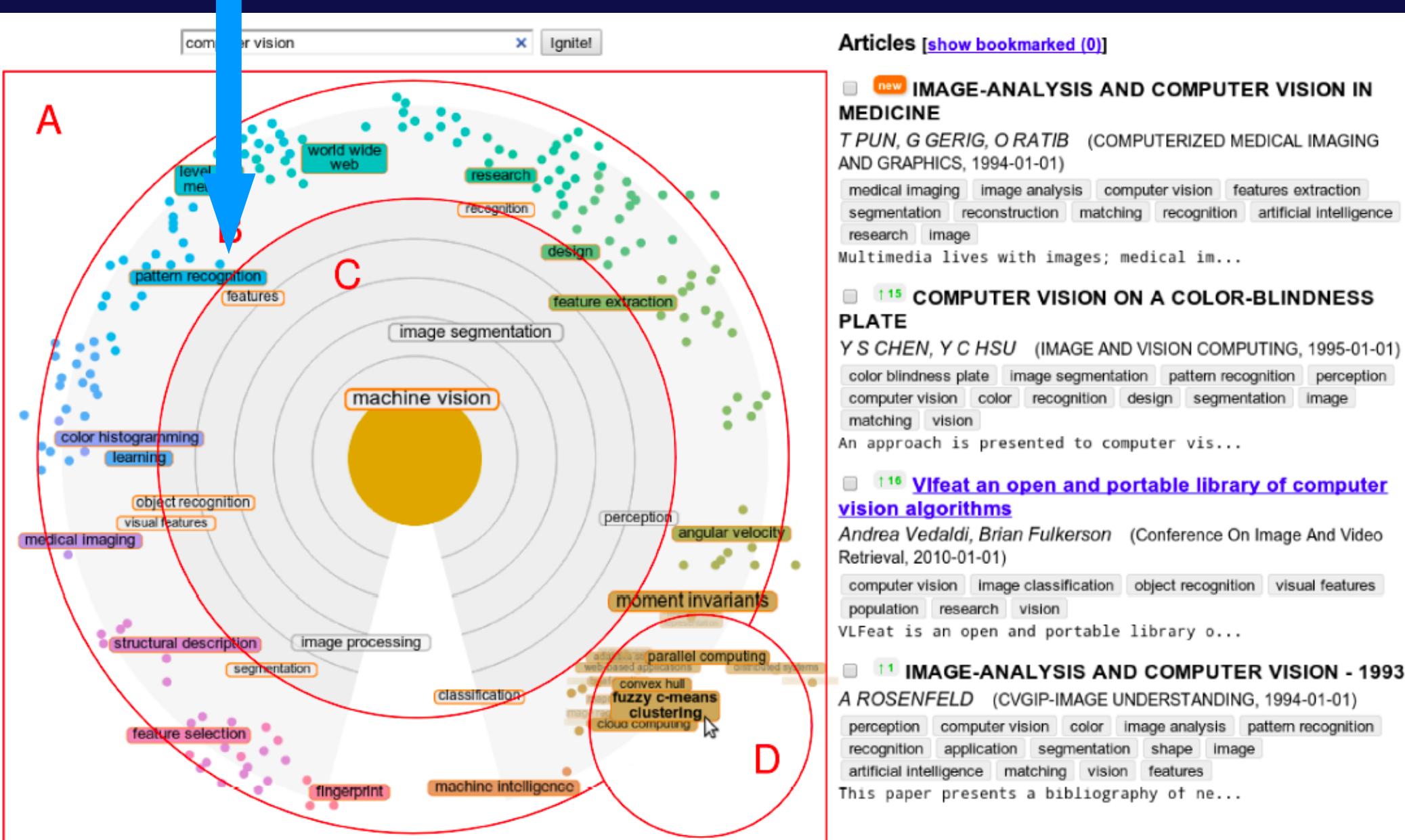
Current intent estimation for which results are retrieved. Angular distance = similarity of intents, radius = relevance

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.

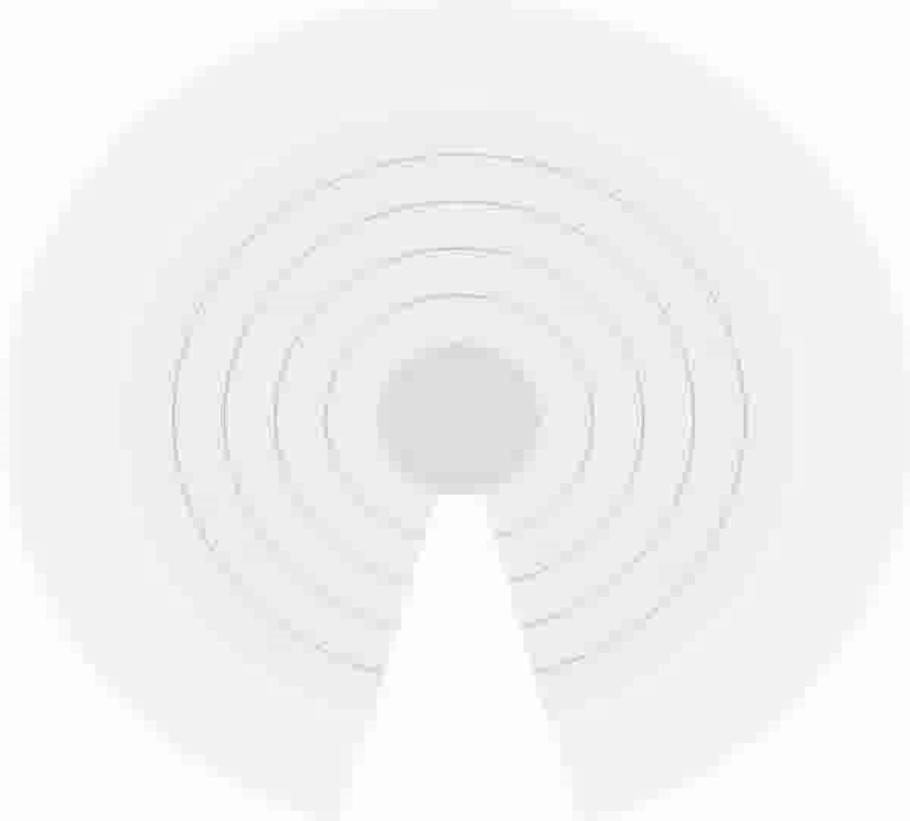


HIIT SCINET

| graphical models|

Ignite

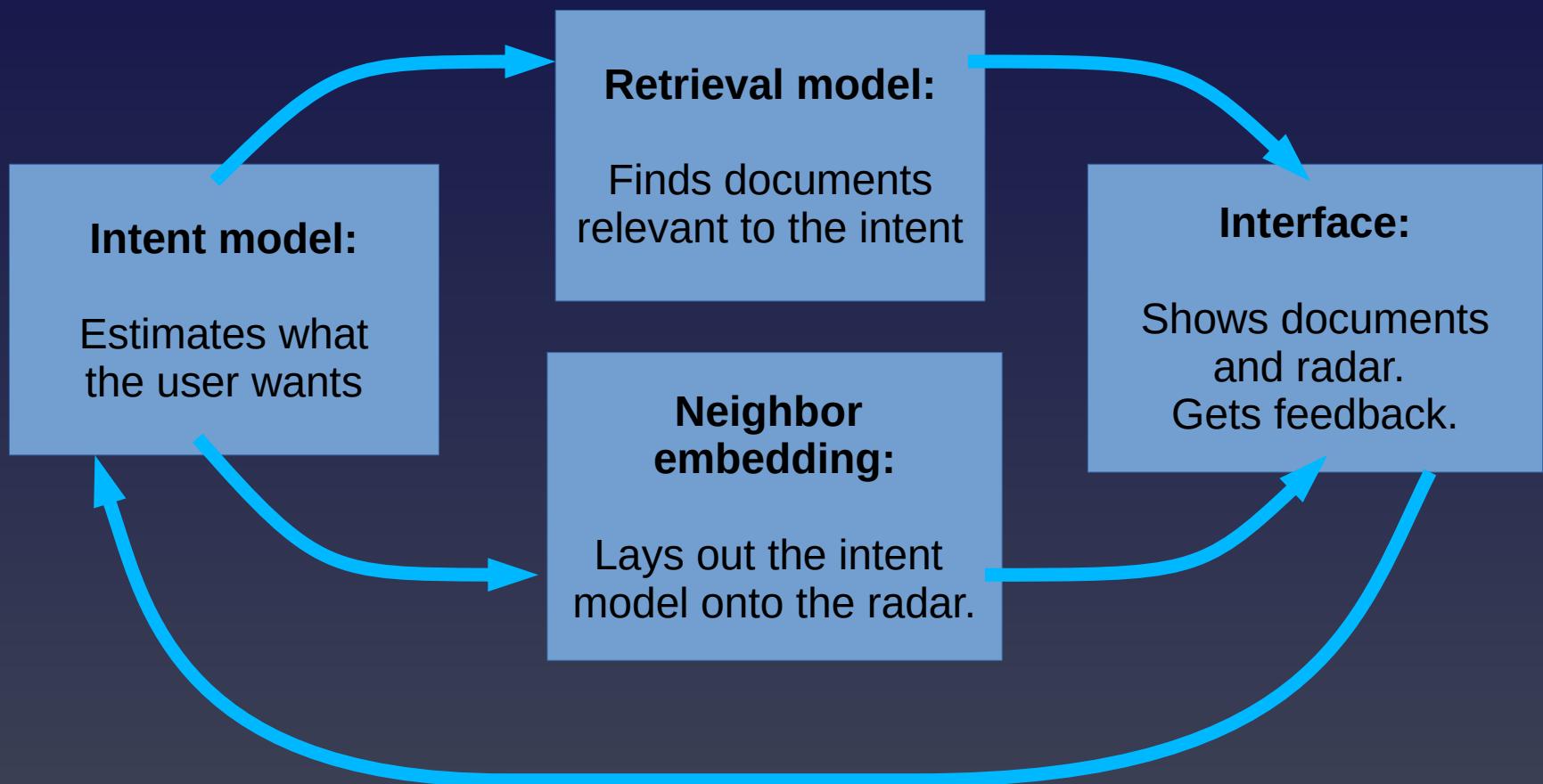
Articles [[show bookmarked \(0\)](#)]



[ALL DONE](#)

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**



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:

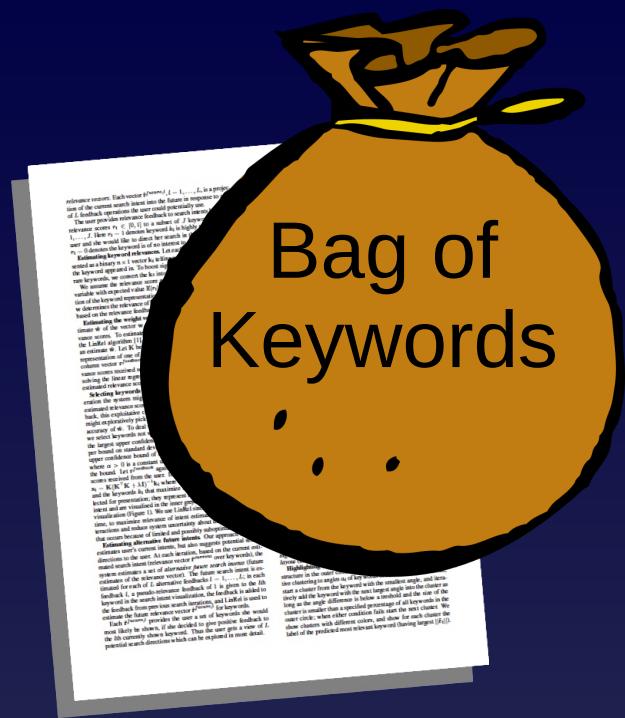
Top positive keyword predictions treated as a sample of an **ideal desired document**

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

$$\begin{aligned} score(d) &= \log P(\mathbf{v}_P | M_d) \\ &= \sum_{i \in K_P} v_i^P \log P(v_i^P | M_d) \end{aligned}$$

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**

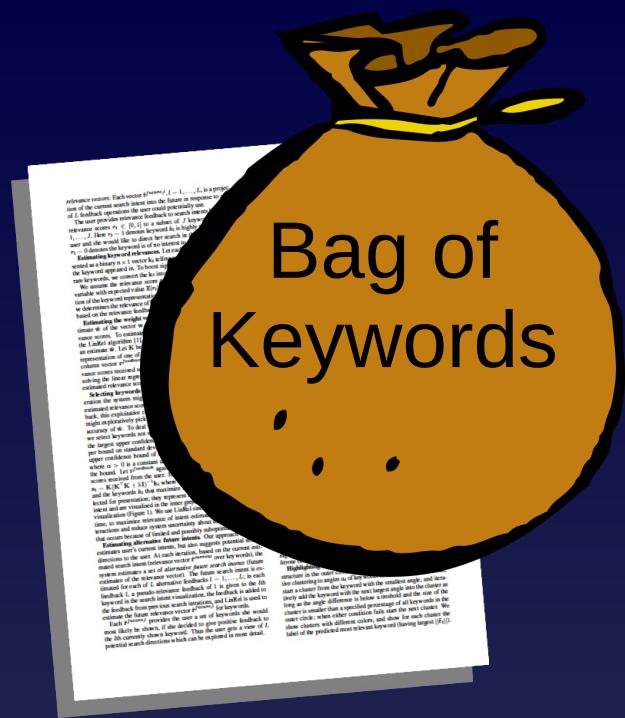
Top negative keyword predictions treated as a sample of an **ideal unwanted document**

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

$$\begin{aligned} score(d) &= \log \frac{P(\mathbf{v}_P | M_d)}{P(\mathbf{v}_N | M_d)} \\ &= \sum_{i \in K_P} v_i^P \log P(v_i^P | M_d) - \sum_{i \in K_N} v_i^N \log P(v_i^N | M_d) \end{aligned}$$

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.



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.

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

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 \mathbf{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.

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

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

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

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

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):

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

Relevance of keyword i
in future intent L

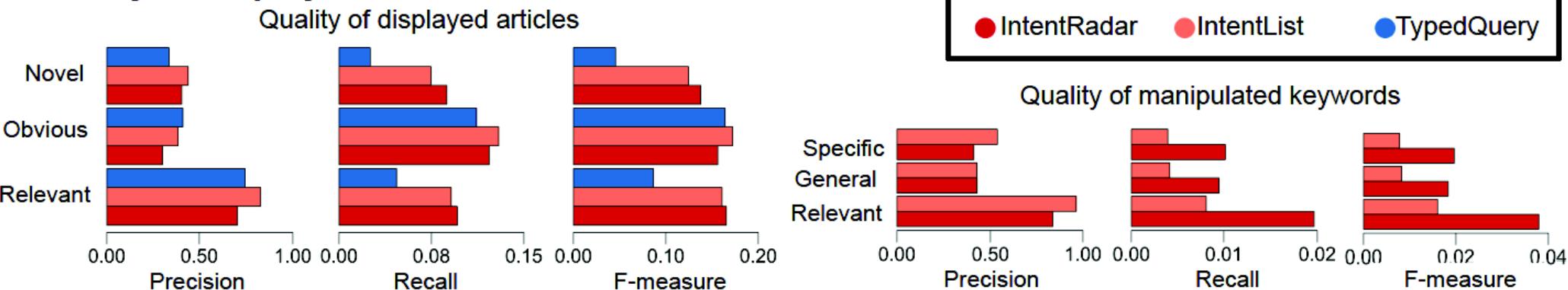
$$\bar{\mathbf{r}}_i = \tilde{\mathbf{r}}_i / \|\tilde{\mathbf{r}}_i\|$$

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

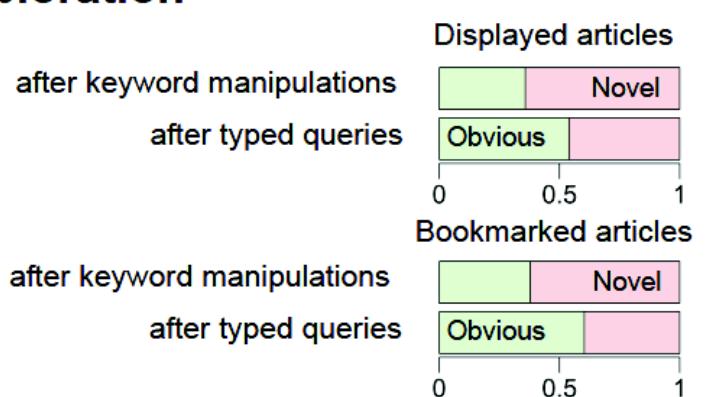
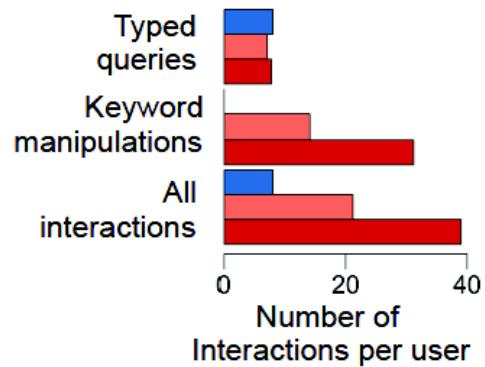
- Layout is **optimized for retrieval of keywords** with similar relevance in future intents, by **nonlinear dimensionality reduction**. We use a well-performing approach optimized for information retrieval, details in Wednesday's talk.

User Experiment - Results

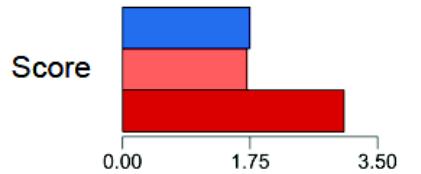
Quality of displayed information



Interaction support for exploration



Task performance



Expert evaluation of written answers of users to their tasks (on a scale 1-5, larger is better)