

Catching up with Industry - Online Evaluation of Information Access Systems

Frank Hopfgartner

@FTHopf



We have a problem!

If SIGIR had an Academic Track, what would be in it?

David Hawking
Microsoft (Bing)
Canberra, Australia
david.hawking@acm.org

ABSTRACT

It used to be the case that very little industry research was presented at SIGIR. Now the balance has radically changed – many accepted papers have industry authors and many rely on industry data sets – To the extent that a leading academic member of the SIGIR community has light-heartedly proposed the creation of an Academic Track.

Behind the levity lies the important question of how a researcher can make a meaningful contribution to the field, in the absence of petabyte-scale sets of documents and massive user-interaction logs. Theoretical contributions can revolutionize thinking, but have greatest impact when applicable in practice, and when empirically validated.

In my years at Funnellback and more recently at Microsoft I have been very aware of high-impact but not-well-solved IR problems involving relatively tiny datasets. Many of them are characterized by sparsity of user interaction data and are hence not well-suited to simple machine learning approaches or to large scale A/B testing. My talk will illustrate and attempt to characterize these problems and to suggest fruitful areas for academic research.

If time permits, I will mention some areas in which academic research has contributed to current large-scale industry practice.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*relevance feedback*

Keywords

Open IR problems for academics.

BIO

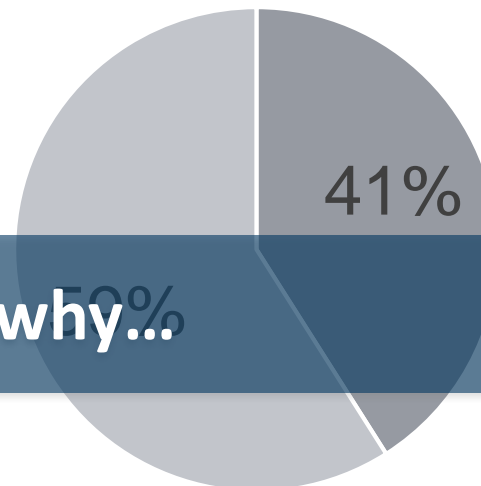
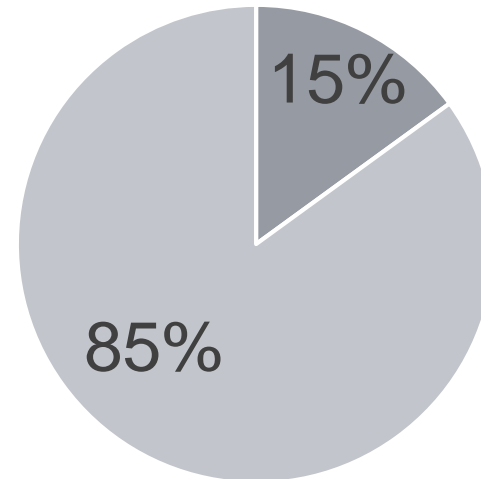
David Hawking is an applied scientist in the Relevance & Intent group of Microsoft. He is based in Canberra and works on Bing.

David has been an IR researcher since 1991 and, among other things, has worked on parallel IR, distributed IR, term proximity models, efficiency, automatic assessment of health webpage quality, evaluation, text collection development, web IR and enterprise search. In 1998 he left the Australian National University to continue his research at Australia's largest research organization CSIRO. CSIRO's industrial focus led to the commercialisation of the team's research and eventually to the spinning off of the Funnellback internet and enterprise search company in late 2005. David joined the spinoff as Chief Scientist in 2008 and worked in that role until joining Microsoft in 2013.

David continues as an adjunct professor at the Australian National University and has graduated nine PhDs. He has authored around a hundred publications in the IR area (see david-hawking.net/publications.shtml) and twice served as program chair for SIGIR. He chaired the Industry Track at SIGIR 2014 and was Web Track coordinator at the International Text Retrieval Conference (TREC) from 1997-2004. In this role he was responsible for the creation and distribution of widely used text retrieval test collections. He holds an honorary doctorate from the University of Neuchâtel (Switzerland), the 2005 Chris Wallace award (Australia) for computer science research and the 2012 Tony Kent Strix award (jointly with Doug Cutting).



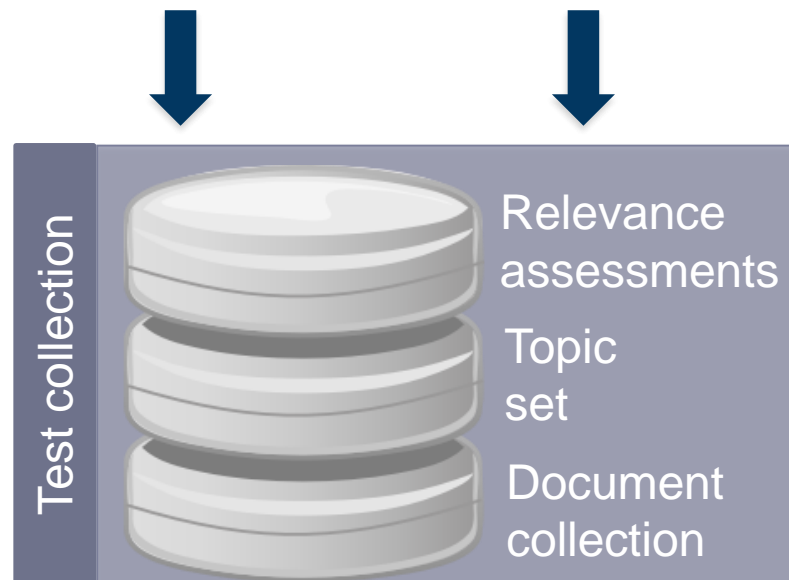
■ Industry ■ Academia



And here is why...

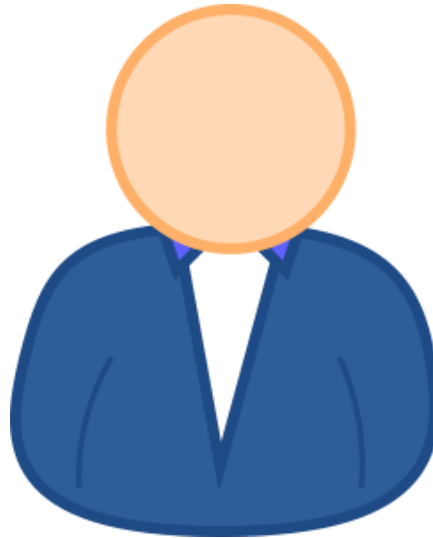
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s). SIGIR '15, August 09 - 13, 2015, Santiago, Chile. ACM 978-1-4503-3621-5/15/08. <http://dx.doi.org/10.1145/2766663.276784>

How do Academics evaluate information access systems?



Academic Evaluation Protocol (Cranfield)

Find as many documents
as possible for a given
search task!

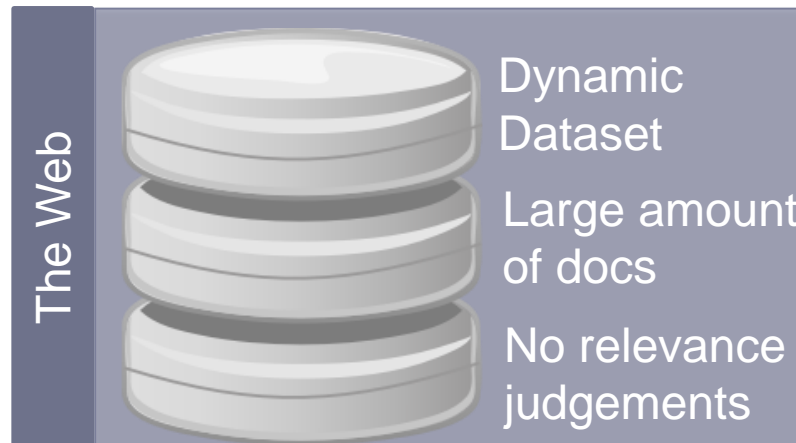
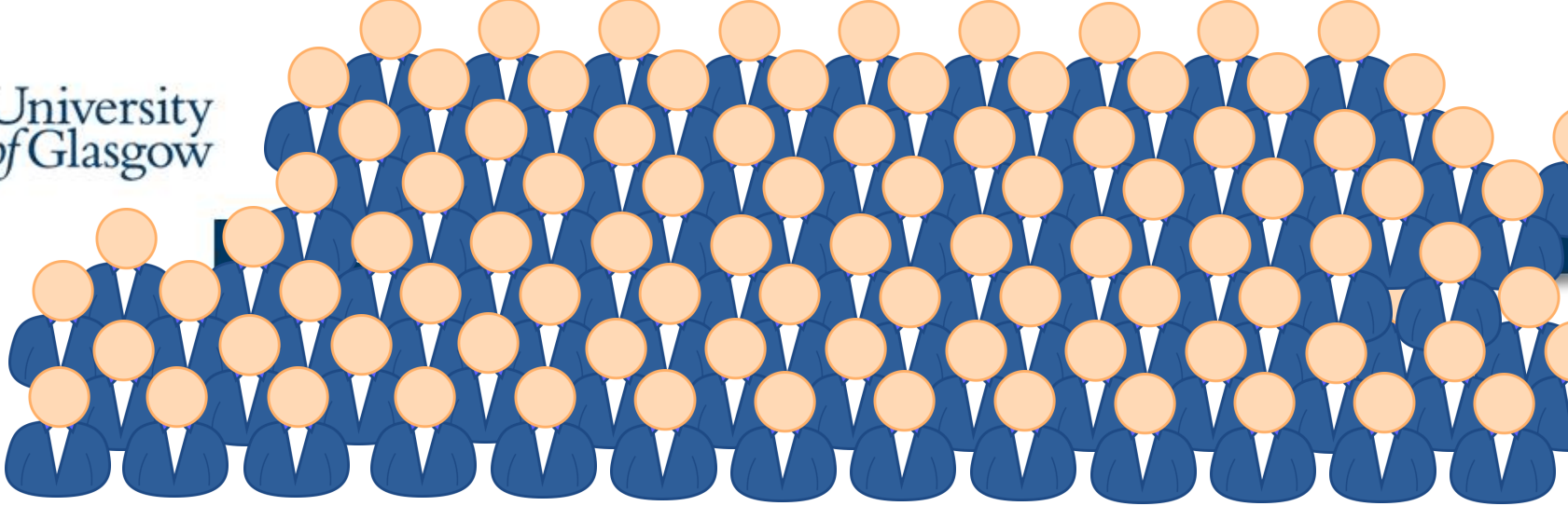


I tell you what is relevant!

Act naturally while I watch
everything you are doing!

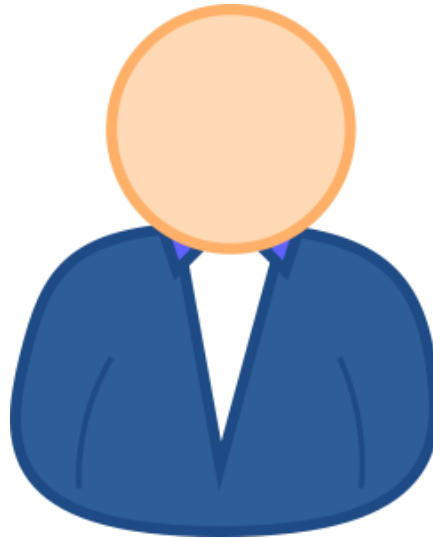


University
of Glasgow



Industry Evaluation Protocol (A/B Testing)

Continue using our
information access
service



Use it for whatever you
want to use it for

Use it whenever and
wherever you want

Ron Kohavi's Keynote at KDD'15

Example 1: MSN Home Page Search Box

- OEC: Clickthrough rate for Search box and popular searches

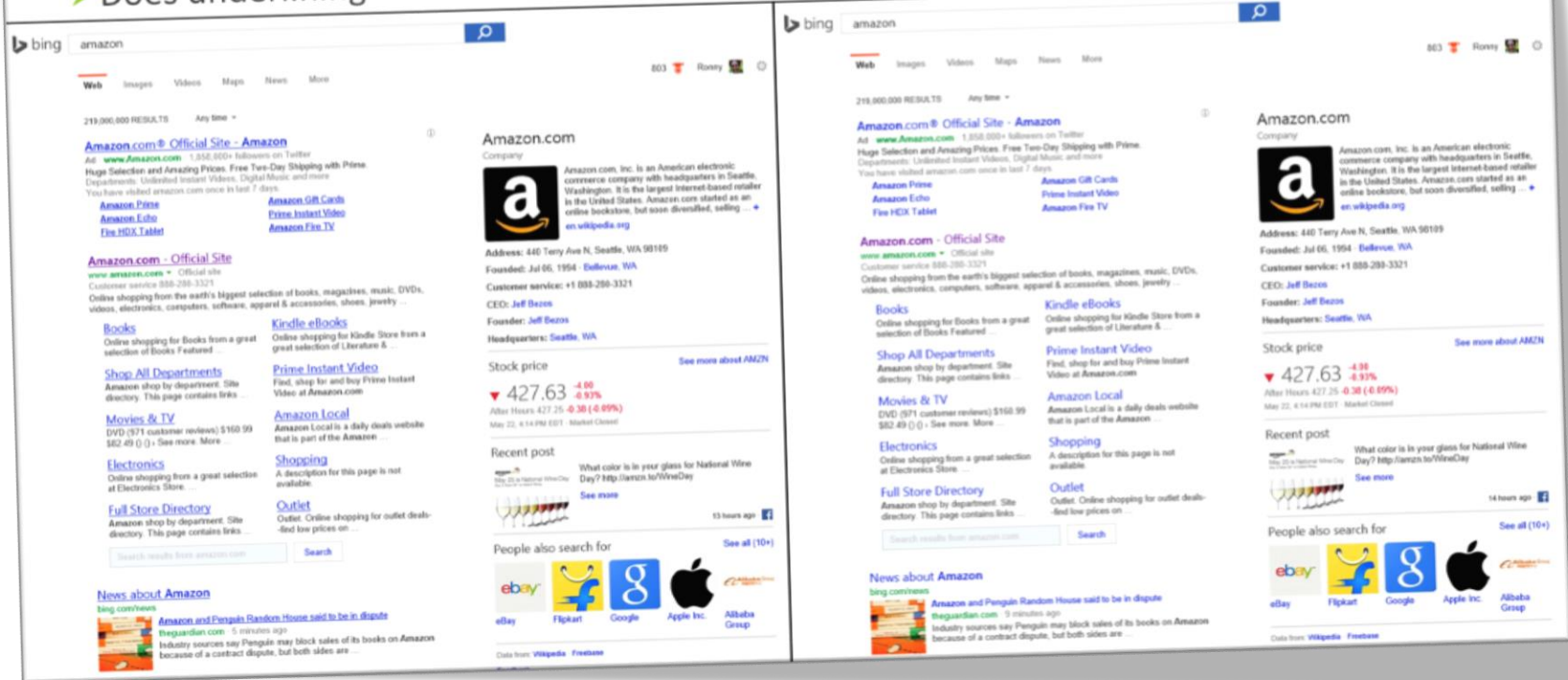


Differences: A has taller search box (overall size is the same),
has magnifying glass icon, "popular searches"
B has big search button, provides popular searches without calling them out

Ron Kohavi's Keynote at KDD'15

Example 4: Underlining Links

➤ Does underlining increase or decrease clickthrough-rate?



The image displays two side-by-side screenshots of a Bing search for 'amazon', illustrating the effect of underlining links on search results. Both screenshots show the same search results, but the left one has underlined links, while the right one does not.

Left Screenshot (Underlined Links):

- Search results for 'amazon' on Bing.
- Top result: [Amazon.com® Official Site - Amazon](#) (underlined).
- Second result: [Amazon.com - Official Site](#) (underlined).
- Third result: [Kindle eBooks](#) (underlined).
- Fourth result: [Prime Instant Video](#) (underlined).
- Fifth result: [Amazon Local](#) (underlined).
- Sixth result: [Shopping](#) (underlined).
- Seventh result: [Outlet](#) (underlined).
- Bottom result: [Full Store Directory](#) (underlined).

Right Screenshot (Non-underlined Links):

- Search results for 'amazon' on Bing.
- Top result: [Amazon.com® Official Site - Amazon](#) (not underlined).
- Second result: [Amazon.com - Official Site](#) (not underlined).
- Third result: [Kindle eBooks](#) (not underlined).
- Fourth result: [Prime Instant Video](#) (not underlined).
- Fifth result: [Amazon Local](#) (not underlined).
- Sixth result: [Shopping](#) (not underlined).
- Seventh result: [Outlet](#) (not underlined).
- Bottom result: [Full Store Directory](#) (not underlined).

Evaluation Gap



Academia

Industry

- Static, often rather old datasets
- Offline Evaluation
- Atypical search task
- Observer-expectancy effect
- Missing context/background
- Missing incentive to satisfy users' information needs

- Dynamic dataset
- Online A/B testing
- Real users
- Real users!
- Real users!!
- Real users!!!

Narrowing the growing gap between Academia and Industry

Do we want **real users** who interact with a system following their own information need?

Do we want **realistic** settings where users are not restricted by closed laboratory conditions?

Do we want to test **scalability** of our systems and algorithms?

Yes

But how do we get access to users and infrastructure?

Option 1: Industry-inclined PhD research

“Doctorates in cooperation with industry do have great value, but must not be seen as a panacea to better prepare doctoral graduates for the job market.”

League of European
Research Universities

Option 2: Living Labs

“A **living laboratory** on the Web that brings researchers and searchers together is needed to facilitate ISSS (Information-Seeking Support System) evaluation.”

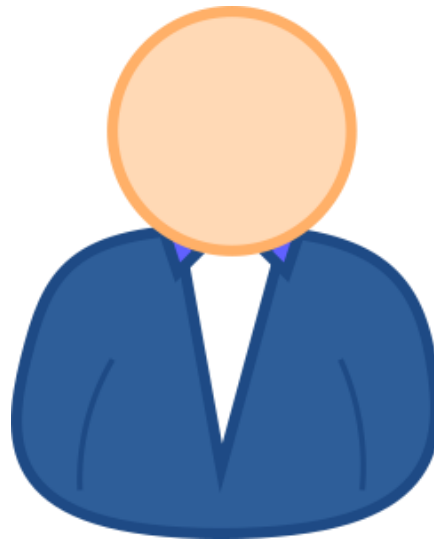
Overview

- Introduction
- **Living Labs**
- News Recommendation Evaluation Lab (NewsREEL) Scenario
- NewsREEL 2016
- NewsREEL and You

What are living labs?

Real-life test and experimentation environment to fill the **pre-commercial gap** between fundamental research and innovation.

Rely on feedback from real users to develop convincing demonstrators that **showcase potentials** of an idea or a product.



Questions to be addressed

Who is
interested in
my product?

Is there a
need for my
product?

How will
people really
use the
technology?

What is the
willingness to
pay?

What
parameters do
I need?

Question

What should a living lab for the promotion of information access research look like?

Search Data is Sensitive!

What Revealing Search Data Reveals

AOL posted, but later removed, a list of the Web search inquiries of 658,000 unnamed users on a new Web site for academic researchers. An interview with one of those unnamed users, Thelma Arnold, combined with her data reveal what she was searching for, why and on which Web sites.

A sample of Thelma Arnold's search data released by AOL

4417749	swing sets	2006-04-24	15:39:30	4	http://www.byoswingset.com
4417749	swing sets	2006-04-24	15:39:30	9	http://www.buychoice.com
4417749	swing sets	2006-04-24	15:39:30	10	http://www.creativeplaythings.com
4417749	swing sets	2006-04-24	15:39:30	5	http://www.childlife.com
4417749	swing sets	2006-04-24	15:39:30	6	http://www.planitplay.com
4417749	that do not shed	2006-04-28	9:05:54	2	http://www.gopetsamerica.com
4417749	dog who urinate on everything	2006-04-28	13:24:07	6	http://www.dogdaysusa.com
4417749	walmart	2006-04-28	14:07:32	1	http://www.walmart.com
4417749	womens underwear	2006-04-28	14:12:28	10	http://www.bizrate.com
4417749	jcpenny	2006-04-28	14:16:05		
4417749	jcpenny	2006-04-28	14:16:49	1	http://www.jcpenny.com
4417749	tortus and turtles	2006-04-29	13:12:47		
4417749	manchester terrier	2006-05-02	9:05:31	1	http://www.manchesterterrier.com
4417749	delta	2006-05-02	11:49:26		
4417749	fingers going numb	2006-05-02	17:35:47		
4417749	dances by laura	2006-05-02	17:59:32		
4417749	dances by lori	2006-05-02	17:59:57		
4417749	single dances	2006-05-02	18:00:18	1	http://solosingles.com
4417749	single dances in atlanta	2006-05-02	18:01:13		
4417749	single dances in atlanta	2006-05-02	18:01:50		
4417749	dry mouth	2006-05-06	16:49:14	2	http://www.mayoclinic.com
4417749	dry mouth	2006-05-06	16:49:14	8	http://www.wrongdiagnosis.com
4417749	thyroid	2006-05-06	16:55:34		
4417749	thyroid	2006-05-06	16:55:44		
4417749	competitive market analysis of homes in lilburn	2006-05-14	12:14:52		
4417749	competitive market analysis of homes in lilburn	2006-05-14	12:16:17		
4417749	competitive market analysis of homes in lilburn	2006-05-14	12:16:43		

Why the search

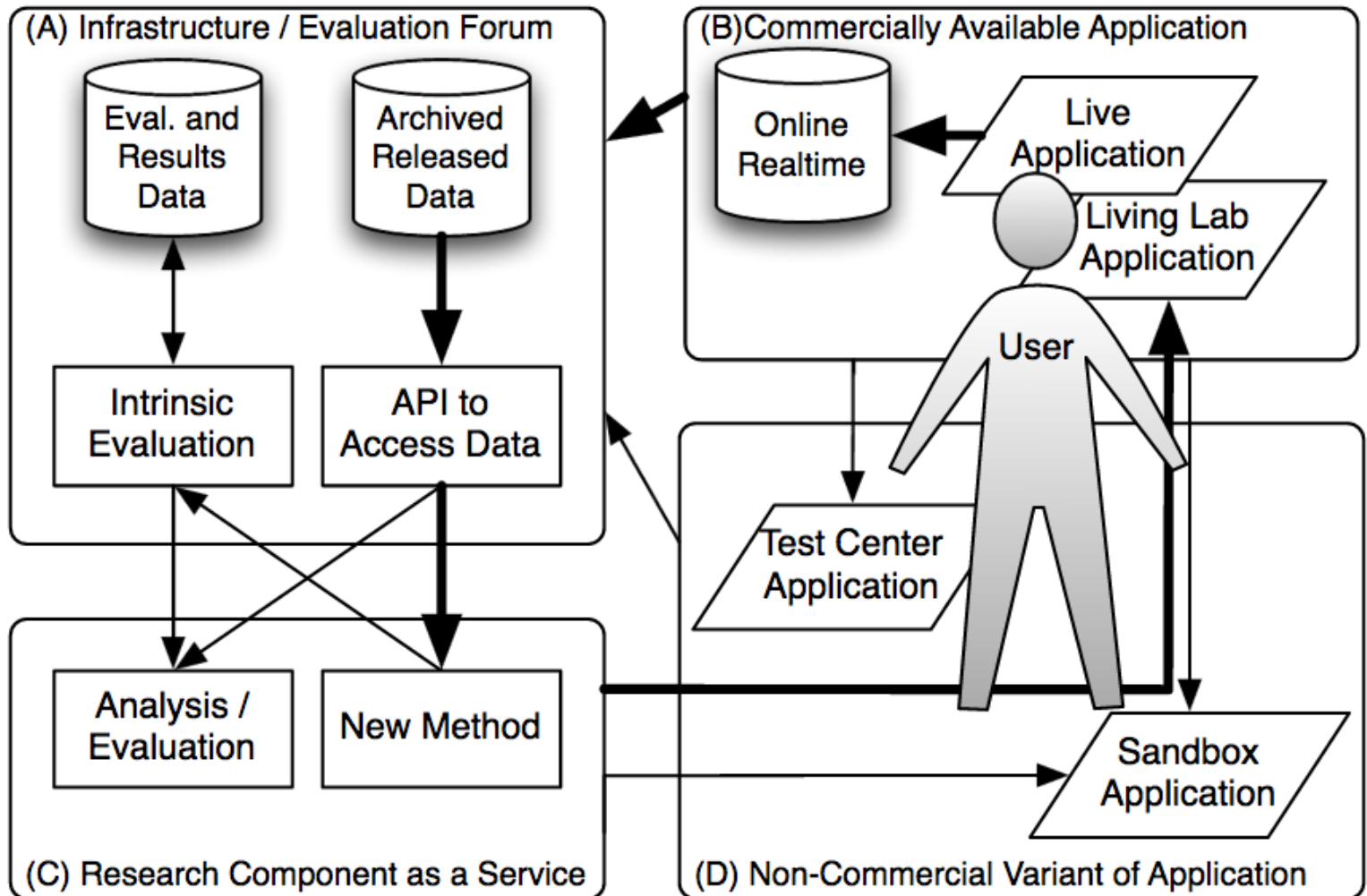
"I was thinking about my grandchildren"

"I was looking for some."

"A woman was in the [public] bathroom crying. She was going through a divorce. I thought there was a place called 'Dances by Lori,' for singles."

"I wanted to find out what my house was worth."

Towards a Living Lab for Information Retrieval...



Living Labs as Academic Research Challenges

- 2013: News Recommender Systems Challenge, organised as workshop and challenge at ACM RecSys'13
- Since 2014: News Recommendation Evaluation Lab (NewsREEL), organised as part of CLEF
- 2016: Living Labs for Information Retrieval (LL4IR), organised as part of CLEF
- Since 2016: OpenSearch, organised as part of TREC
- Since 2016: Open Live Test for Question Retrieval, organised as part of NTCIR

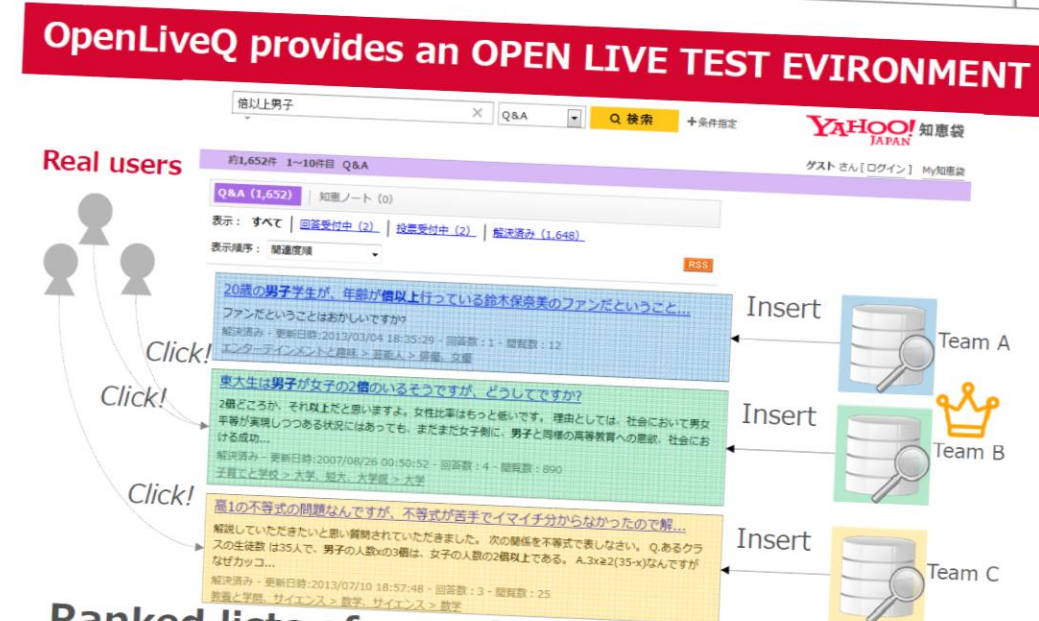
TREC OpenSearch 2016

KNOWN ISSUES

- Head queries only
 - Considerable portion of traffic, but only popular info needs
- Lack of context
 - No knowledge of the searcher's location, previous searches, etc.
- No real-time feedback
 - API provides detailed feedback, but it's not immediate
- Limited control
 - Experimentation is limited to single searches, where results are interleaved with those of the production system; no control over the entire result list
- Ultimate measure of success
 - Search is only a means to an end, it is not the ultimate goal

NTCIR-13: Open Live Test for Question Retrieval

OpenLiveQ provides an OPEN LIVE TEST ENVIRONMENT



Real users

Click!

Click!

Click!

Insert

Team A

Insert

Team B

Insert

Team C

Ranked lists of questions from participants' systems are **INTERLEAVED, presented to real users, and evaluated by their clicks**



News REcommendation Evaluation Lab (NewsREEL)

Lab Overview

Frank Hopfgartner, Benjamin Kille, Andreas Lommatzsch, Martha Larson, Torben Brodt, Jonas Seiler



Overview

- Introduction
- Living Labs
- **News Recommendation Evaluation Lab (NewsREEL) Scenario**
- NewsREEL 2016
- NewsREEL and You

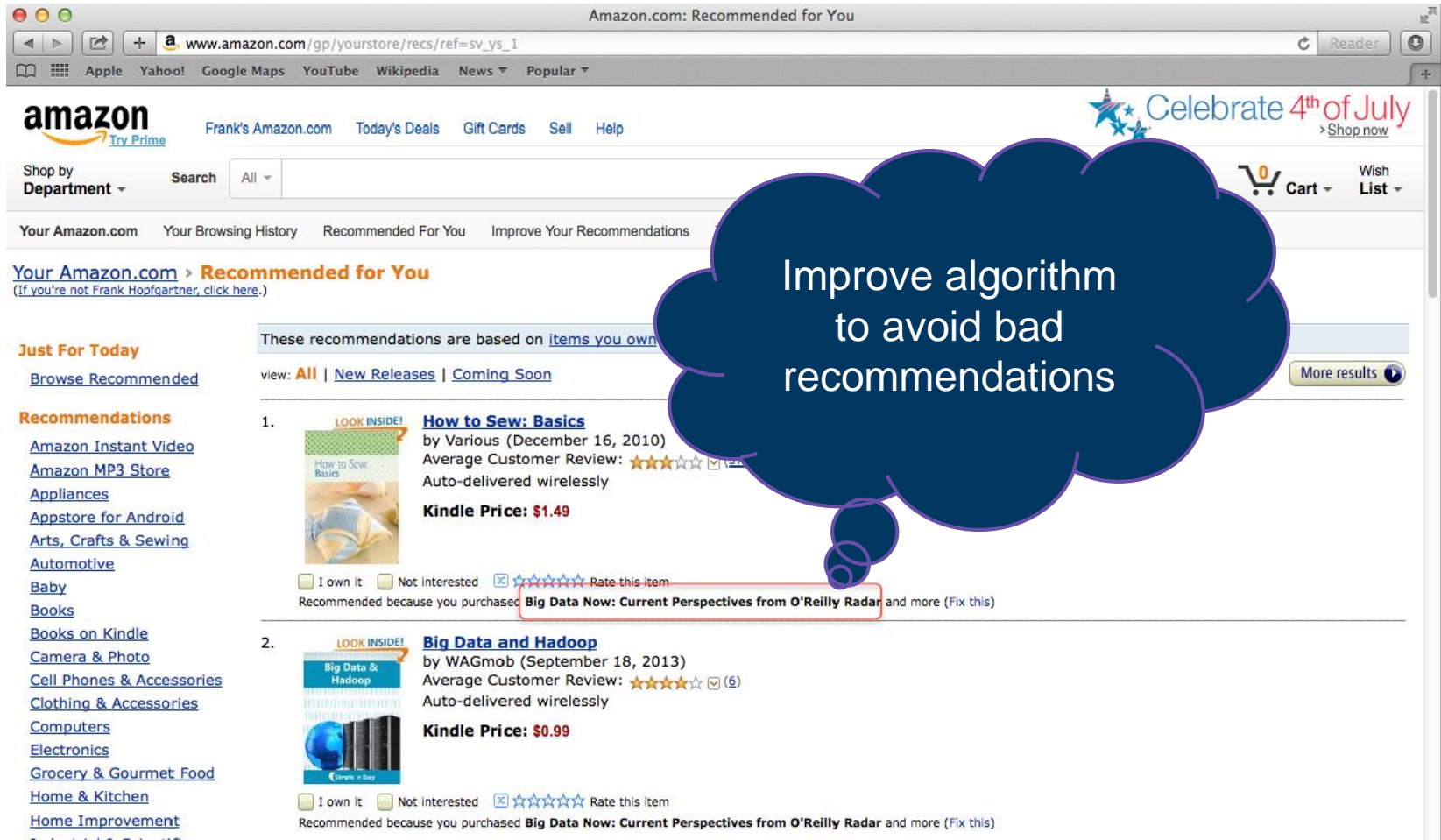
CLEF NewsREEL

In CLEF NewsREEL, participants can develop stream-based **news recommendation** algorithms and have them benchmarked (a) online by **millions of users** over the period of a few months, and (b) offline by **simulating a live stream**.

What are recommender systems?

Recommender systems help users to find **items** they were not searching for.

Example: Online Retailer

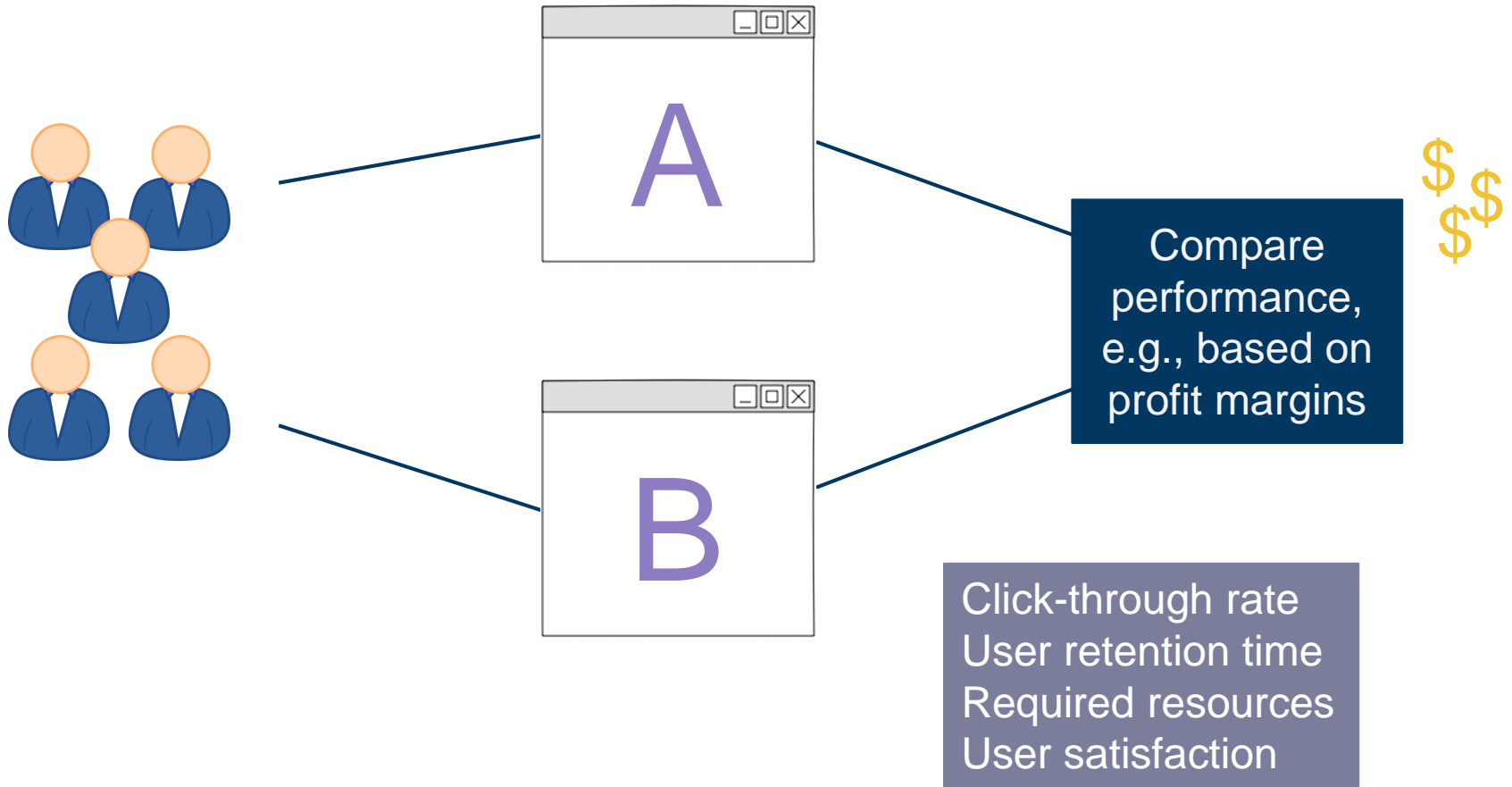


The screenshot shows the Amazon.com interface for a user named Frank. The page title is "Amazon.com: Recommended for You". The URL is www.amazon.com/gp/yourstore/recs/ref=sv_ys_1. The page features a search bar, navigation links, and a "Celebrate 4th of July" banner. The main content area is titled "Recommended for You" and displays two book recommendations:

- How to Sew: Basics** by Various (December 16, 2010). Average Customer Review: 4.5 stars. Kindle Price: \$1.49. Recommended because you purchased **Big Data Now: Current Perspectives from O'Reilly Radar** and more.
- Big Data and Hadoop** by WAGmob (September 18, 2013). Average Customer Review: 4.5 stars. Kindle Price: \$0.99. Recommended because you purchased **Big Data Now: Current Perspectives from O'Reilly Radar** and more.

A large blue thought bubble with a purple outline is overlaid on the right side of the page, containing the text: "Improve algorithm to avoid bad recommendations". A red box highlights the text "Recommended because you purchased Big Data Now: Current Perspectives from O'Reilly Radar" under both book listings.

Online Evaluation (A/B testing)




NewsREEL scenario

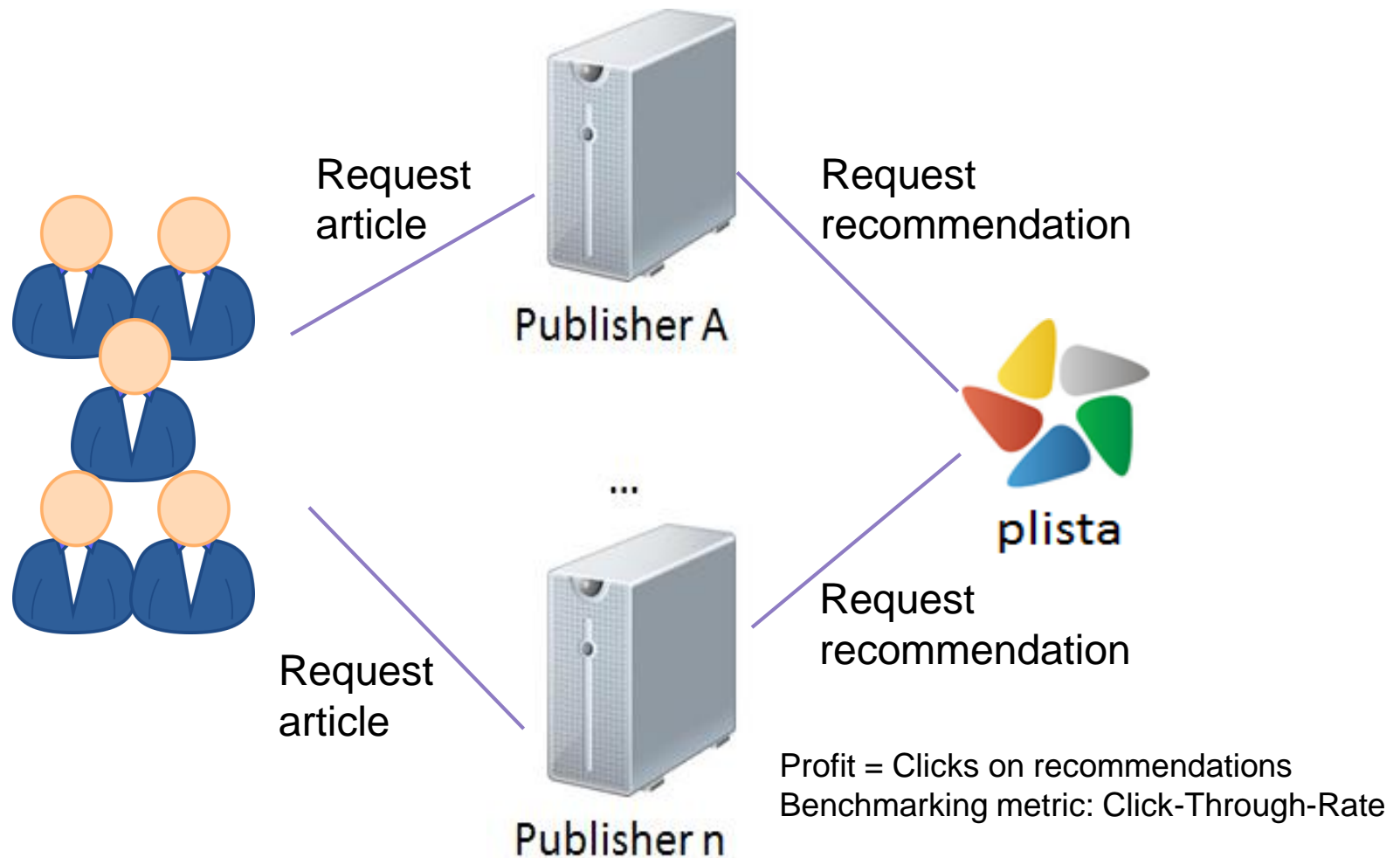


Das könnte Sie auch interessieren

- 
Daran erkennt man gute Autovermietungen
 Eine Autovermietung kann während des Urlaubs sehr nützlich sein. Doch woran erkennt man, dass dr Autovermieter seriös ist?... mehr
- 
VW Jetta
 Der Jetta kehrt zurück nach Europa. Der sportlicher denn je konzipierte Volkswagen soll nun das Limousinen-Spektrum komplettieren. Vergleichen Sie... mehr
- 
Was bringt ein schadstoffarmer PKW
 Umweltbewusstsein ist gefragt, daher setzen immer mehr Automobilhersteller auf schadstoffarme PKW. Doch viele Verbraucher zögern beim Kauf ... mehr
- 
Beim Autokauf ein Schnäppchen machen
 Statistisch gesehen kauft sich jeder Deutsche in seinem Leben fast 11 Autos. Da der Kauf eines Autos nicht gerade günstiges Unterfangen ist, mehr

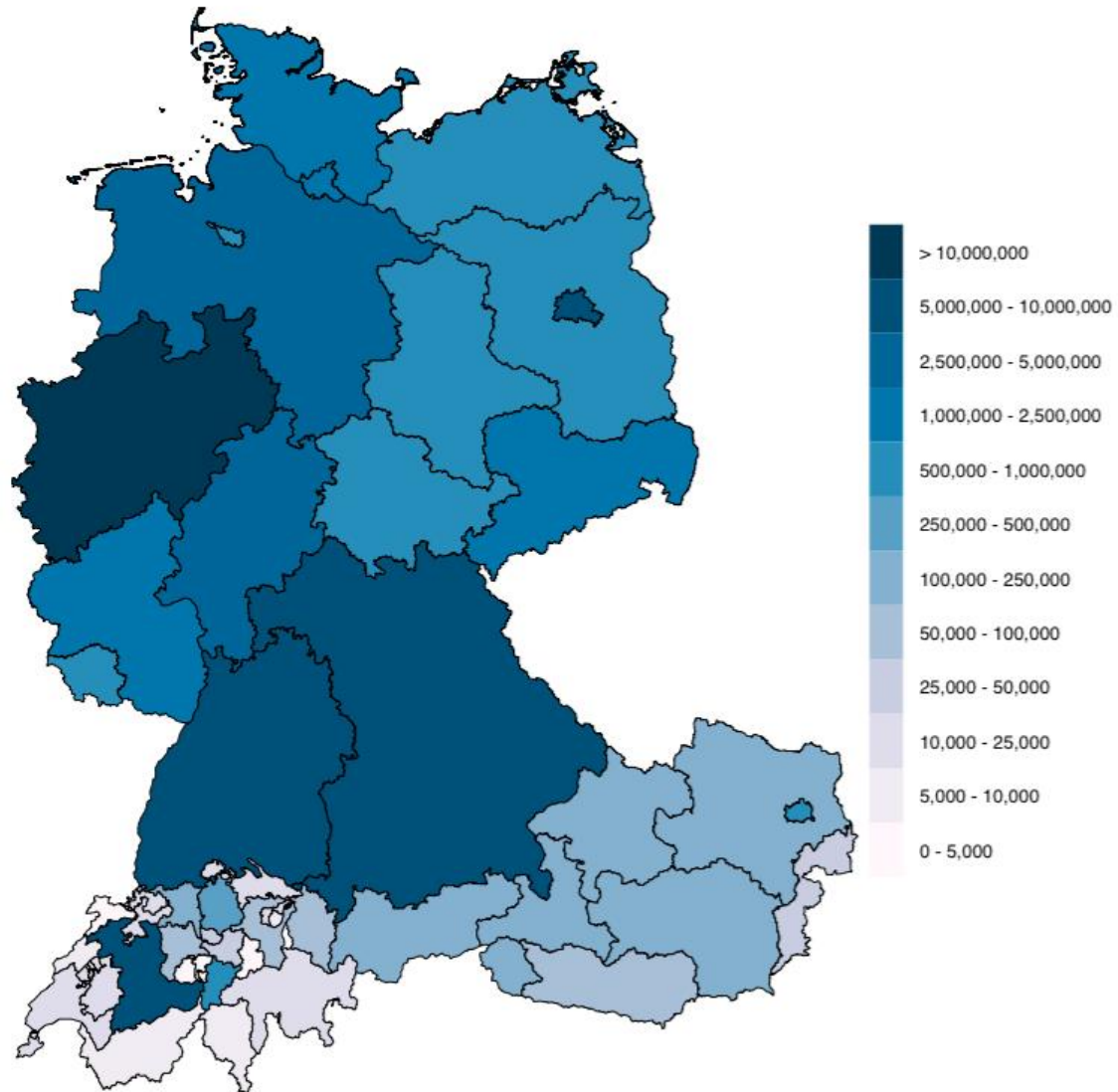
hier werben 

NewsREEL scenario



Who are the users?

>80%



Task 1: Online Evaluation

Recommend news articles in real-time

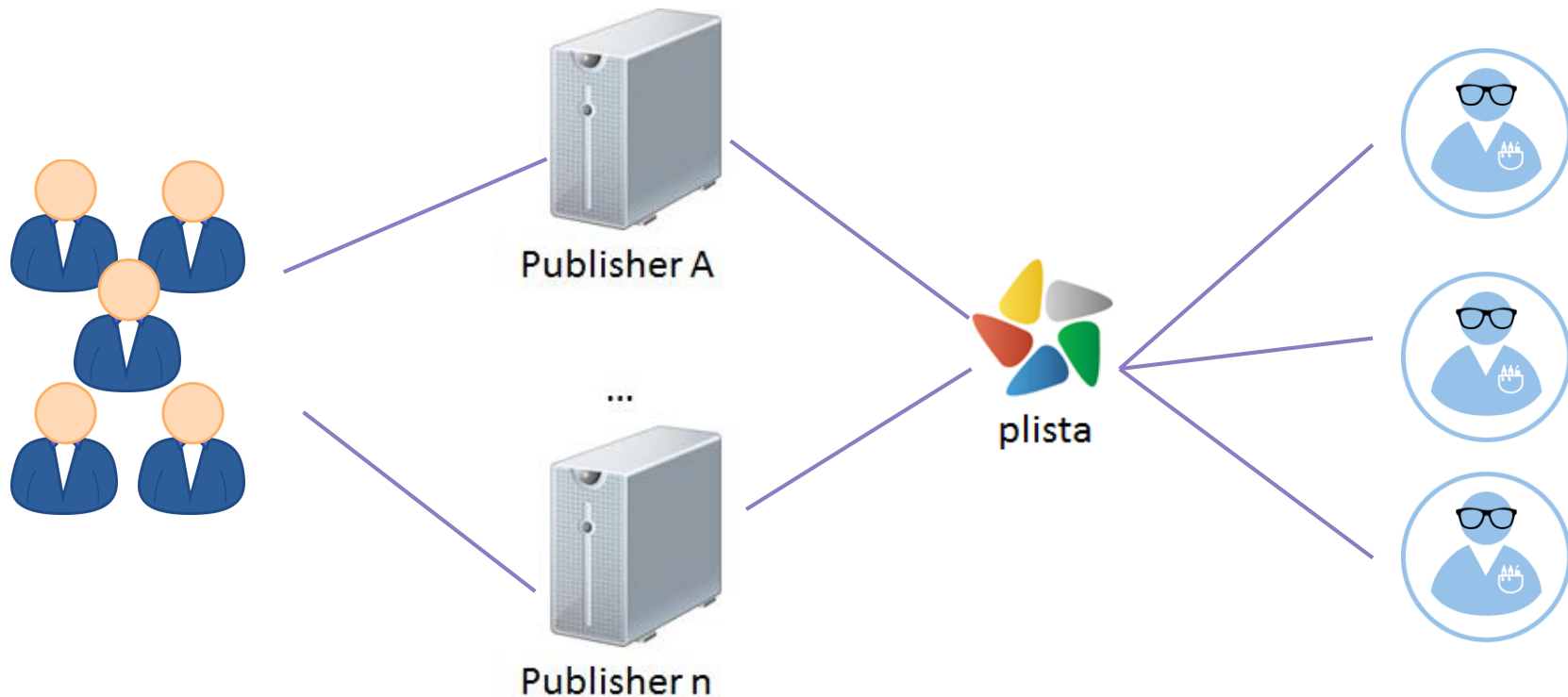
Data

- Provide recommendations for visitors of the news portals of plista's customers
- Ten portals (local news, sports, business, technology)
- Communication via [Open Recommendation Platform \(ORP\)](#)

Evaluation

- Benchmark own performance with other participants and baseline algorithms during three pre-defined evaluation windows
- Best algorithms determined in final evaluation period
- Standard evaluation metrics

Real-Time Recommendation



CLEF-NEWSREEL

NEWS RECOMMENDATION EVALUATION LAB



HOME TASKS HOW TO PARTICIPATE PUBLICATIONS ORGANISATION PREVIOUS CAMPAIGNS ✕ CLEF 2017 PROGRAM

Overview

Commercial providers of information access systems (such as Amazon or Google) usually evaluate the performance of their algorithms by observing how large numbers of their customers interact with different instances of their services. Unfortunately, due to the lack of access to large-scale systems, university-based research is struggling to catch up with this large-scale evaluation methodology. NewsREEL, short for News Recommendation Evaluation Lab, aims to bridge this "evaluation gap" between Academia and Industry.

NewsREEL is organised as a campaign-style evaluation lab of CLEF 2017 and addresses the following information access task:

Whenever a visitor of an online news portal reads a news article on their side, the task is to recommend other news articles that the user might be interested in.

NewsREEL offers two tasks to study this use case. The first task, **NewsREEL Live**, implements the idea of a "living lab" where the provider of a recommendation service provides access to its infrastructure and user base. The second task, **NewsREEL Replay**, replays a live setting using the recommender system reference framework **Idomaar**.

By providing this service for millions of users, the recommendation scenario requires solutions to significant research challenges, such as processing information in real-time, handling vast amount of data, and providing suitable



Important Dates

- Labs registration opens: 4 November 2016
- Registration Closes: 21 April 2017
- Test Period 1: TBA
- Test Period 2: TBA
- Test Period 3: TBA
- Evaluation Period: TBA
- End Evaluation Cycle: 5 May 2017
- Submission of Working Notes: 26 May 2017
- Feedback on Working Notes: 16 June 2017
- Camera Ready Working Notes due: 3 July 2017
- CLEF 2017: 11-14 September 2017

@clefnewsreel

The poster is on "The Potentials of Recommender Systems Challenges for Student Learning"

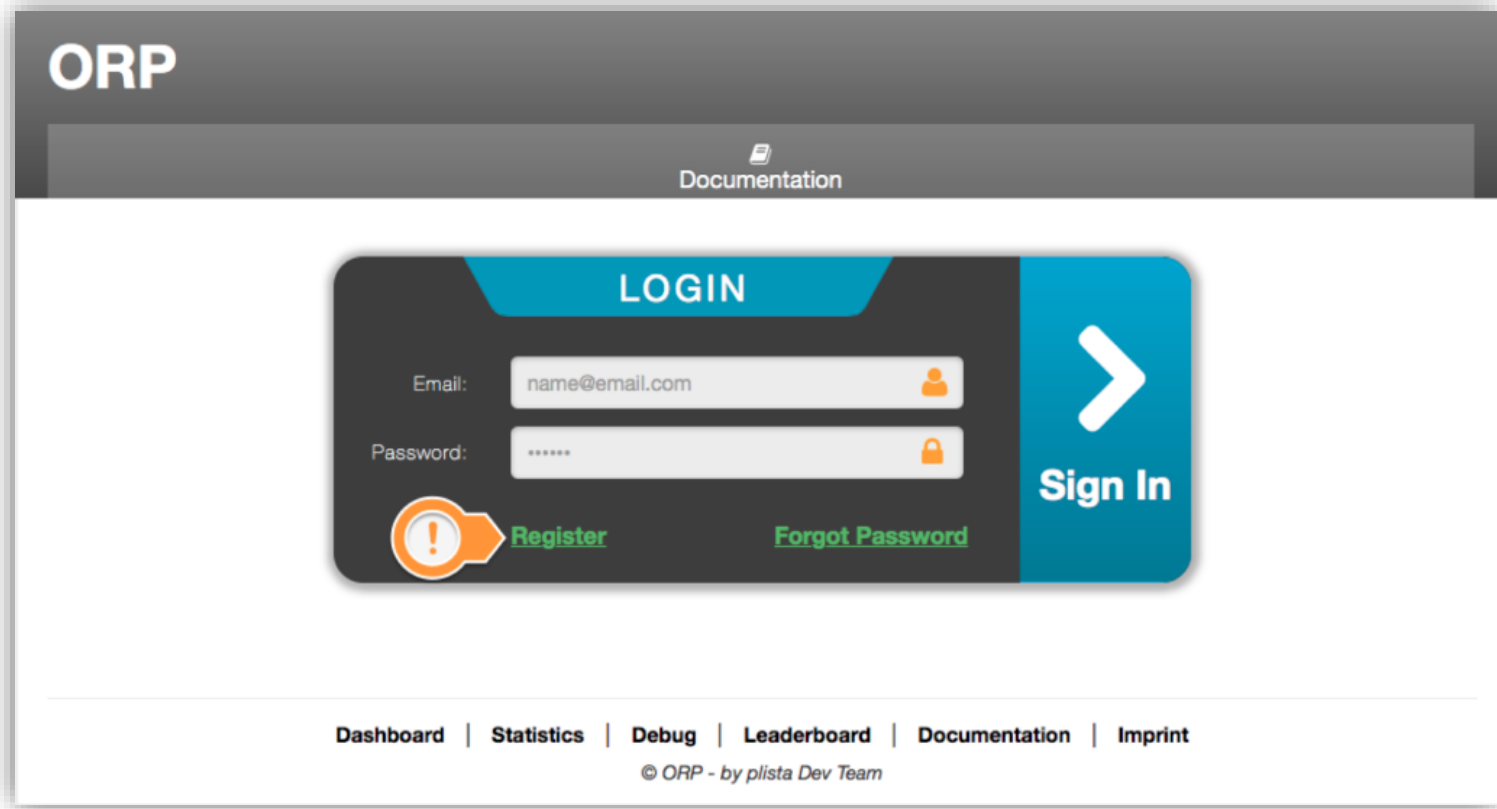
Getting Started: Set up a server

- There are different implementations which handle ORP

- PHP
- Java
- Node.js

```
-- orp-sdk-java
|-- log4j.properties
|-- pom.xml
|-- README.md
|-- src
    |-- main
        |-- java
            |-- de
                |-- dailab
                    |-- plistacontest
                        |-- client
                            |-- Client.java
                            |-- ContestHandler.java
                            |-- DirtyRingBuffer.java
                            |-- RecommenderItem.java
                            |-- RecommenderItemTable.java
        |-- resource
            |-- log4j.properties
            |-- log4j.xml
            |-- plistacontest.properties
9 directories, 11 files
```

Getting Started: Register (AFTER registering via CLEF)



The screenshot shows the ORP (Open Research Platform) interface. At the top left, the text "ORP" is displayed in a large, bold, white font on a dark grey background. Below this, a dark grey navigation bar contains a document icon and the word "Documentation". The main content area is white and features a dark grey login/register card. The card has a teal header with the word "LOGIN" in white. Below the header, there are two input fields: "Email:" with the placeholder "name@email.com" and a user icon, and "Password:" with a masked password "*****" and a lock icon. To the right of these fields is a large teal button with a white right-pointing arrow and the text "Sign In". Below the input fields, there are two links: "Register" with a warning icon and "Forgot Password". At the bottom of the page, a white navigation bar contains links for "Dashboard", "Statistics", "Debug", "Leaderboard", "Documentation", and "Imprint". Below the navigation bar, the copyright notice "© ORP - by plista Dev Team" is displayed.


ORP

Documentation

LOGIN

Email:

Password:

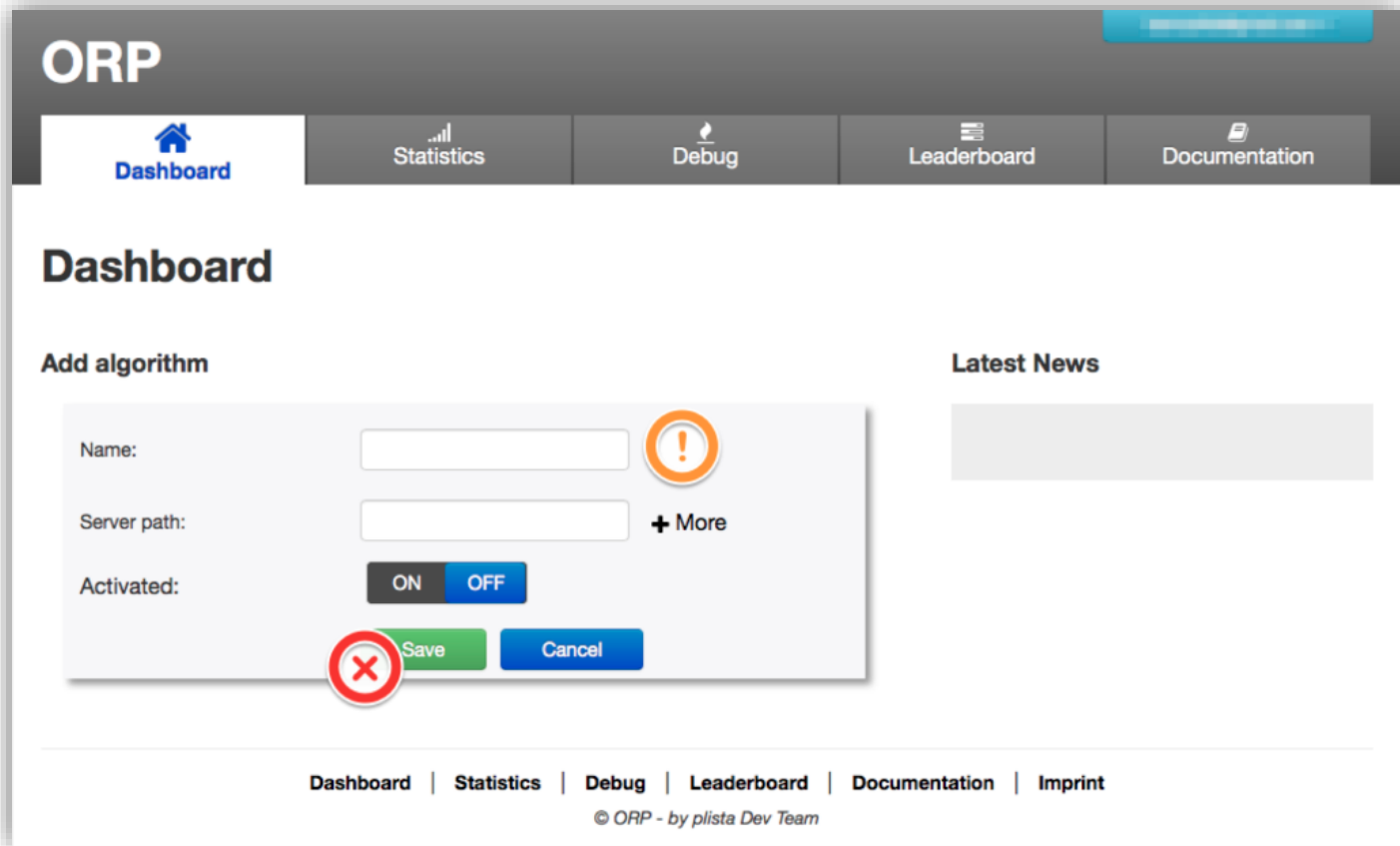
 [Register](#) [Forgot Password](#)

Sign In

[Dashboard](#) | [Statistics](#) | [Debug](#) | [Leaderboard](#) | [Documentation](#) | [Imprint](#)

© ORP - by plista Dev Team

Getting Started: Log into Open Recommender Platform




The screenshot displays the ORP Dashboard interface. At the top, the 'ORP' logo is visible on the left, and a blue button is on the right. Below the logo, a navigation bar contains five items: 'Dashboard' (with a home icon), 'Statistics' (with a signal icon), 'Debug' (with a flame icon), 'Leaderboard' (with a list icon), and 'Documentation' (with a document icon). The main content area is titled 'Dashboard' and is divided into two sections: 'Add algorithm' and 'Latest News'. The 'Add algorithm' section features a form with three input fields: 'Name:', 'Server path:', and 'Activated:'. The 'Activated:' field has 'ON' and 'OFF' toggle buttons. Below the form are 'Save' and 'Cancel' buttons. A red 'X' icon is positioned over the 'Save' button. An orange warning icon is located to the right of the 'Name:' field, and a '+ More' link is to the right of the 'Server path:' field. The 'Latest News' section is currently empty. At the bottom of the page, a footer contains navigation links: 'Dashboard | Statistics | Debug | Leaderboard | Documentation | Imprint' and the copyright notice '© ORP - by plista Dev Team'.

ORP

Dashboard | Statistics | Debug | Leaderboard | Documentation


Dashboard

Add algorithm

Name: 

Server path: + More

Activated:



Latest News

Dashboard | Statistics | Debug | Leaderboard | Documentation | Imprint

© ORP - by plista Dev Team

ORP

Dashboard Statistics Debug Leaderboard Documentation

performance monitor compare to others docs

Dashboard

RinginBuff

0	Requests	▲ 0
0	Clicks	✖
0%	CTR	

* Statistics shown represent data for the past five (5) days

✓

+ Add new algorithm

Algorithm Test

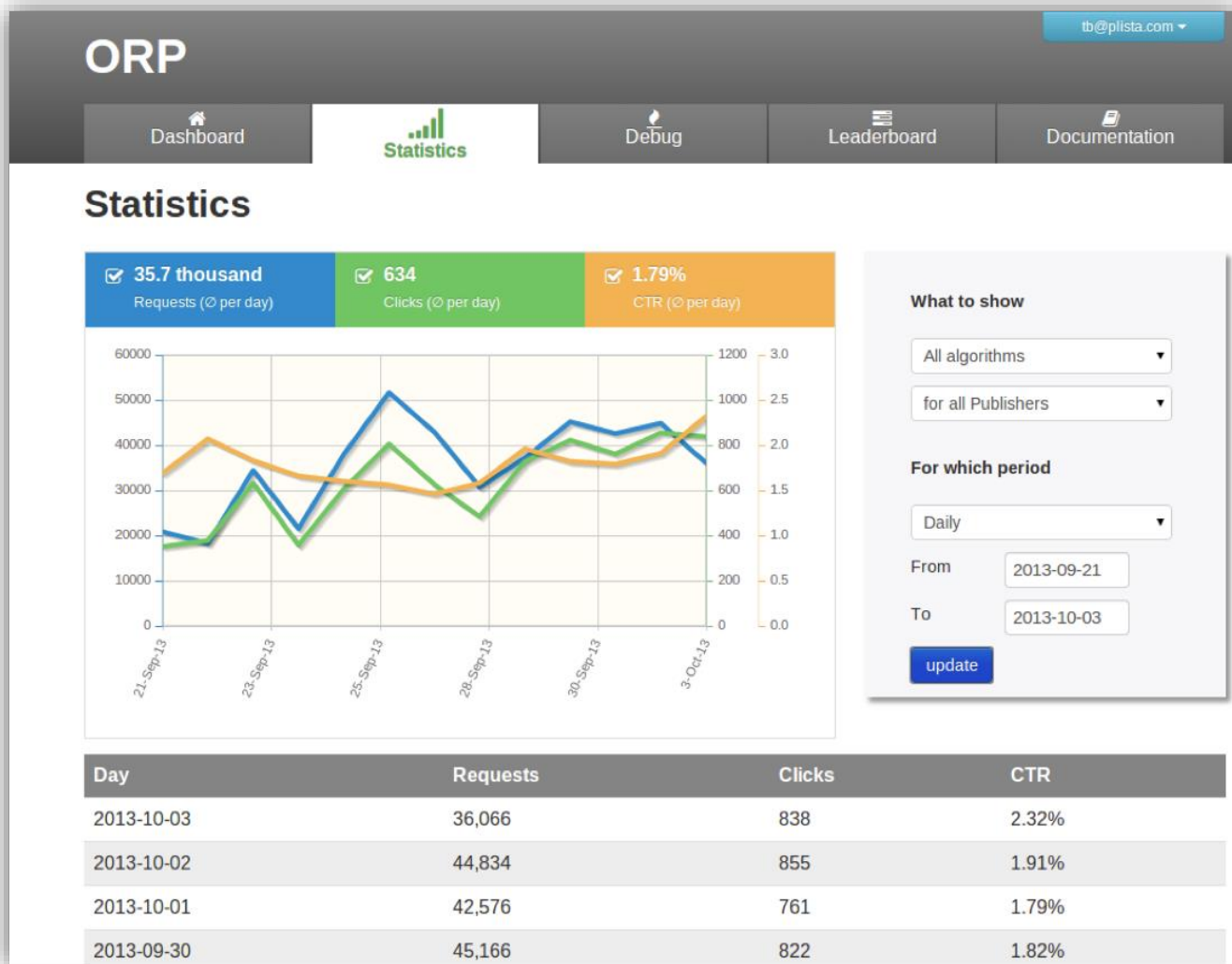
Algorithm: RinginBuff

Trigger-Type: Requests

Fire

Latest News

Getting Started: Monitor performance



Overview

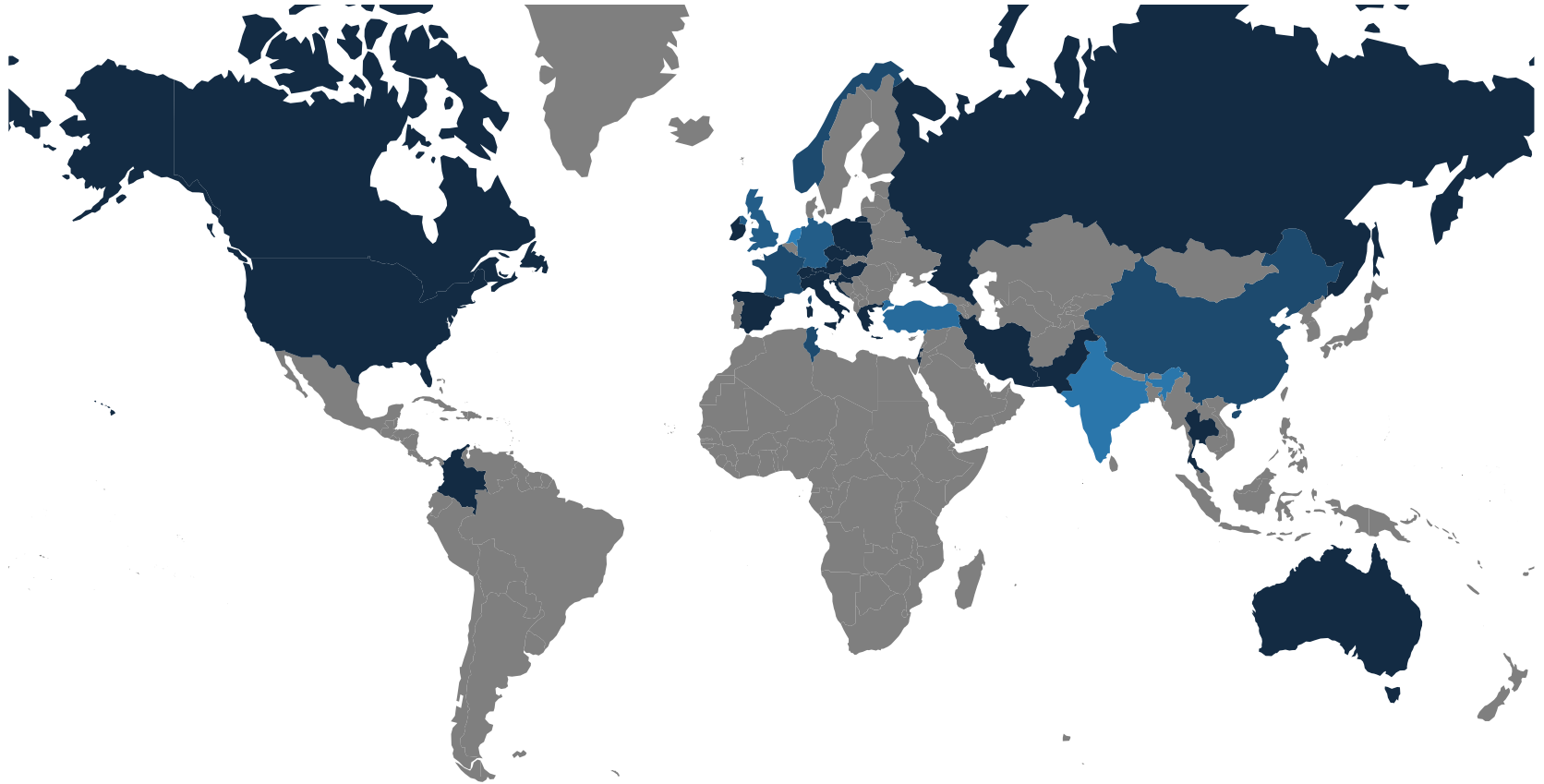
- Introduction
- Living Labs
- News Recommendation Evaluation Lab (NewsREEL) Scenario
- **NewsREEL 2016**
- NewsREEL and You

Promotion activities

- Call for participations
- Social media and website
- Tutorials at ACM RecSys'15, ECIR'15, workshop presentations
- Invited talks at workshops, institutes, meetup groups
- Demo at ACM RecSys'16
- ...



Registrations for NewsREEL 2016



CLEF NewsREEL 2016 Participation



Evaluation Schedule (Task 1)

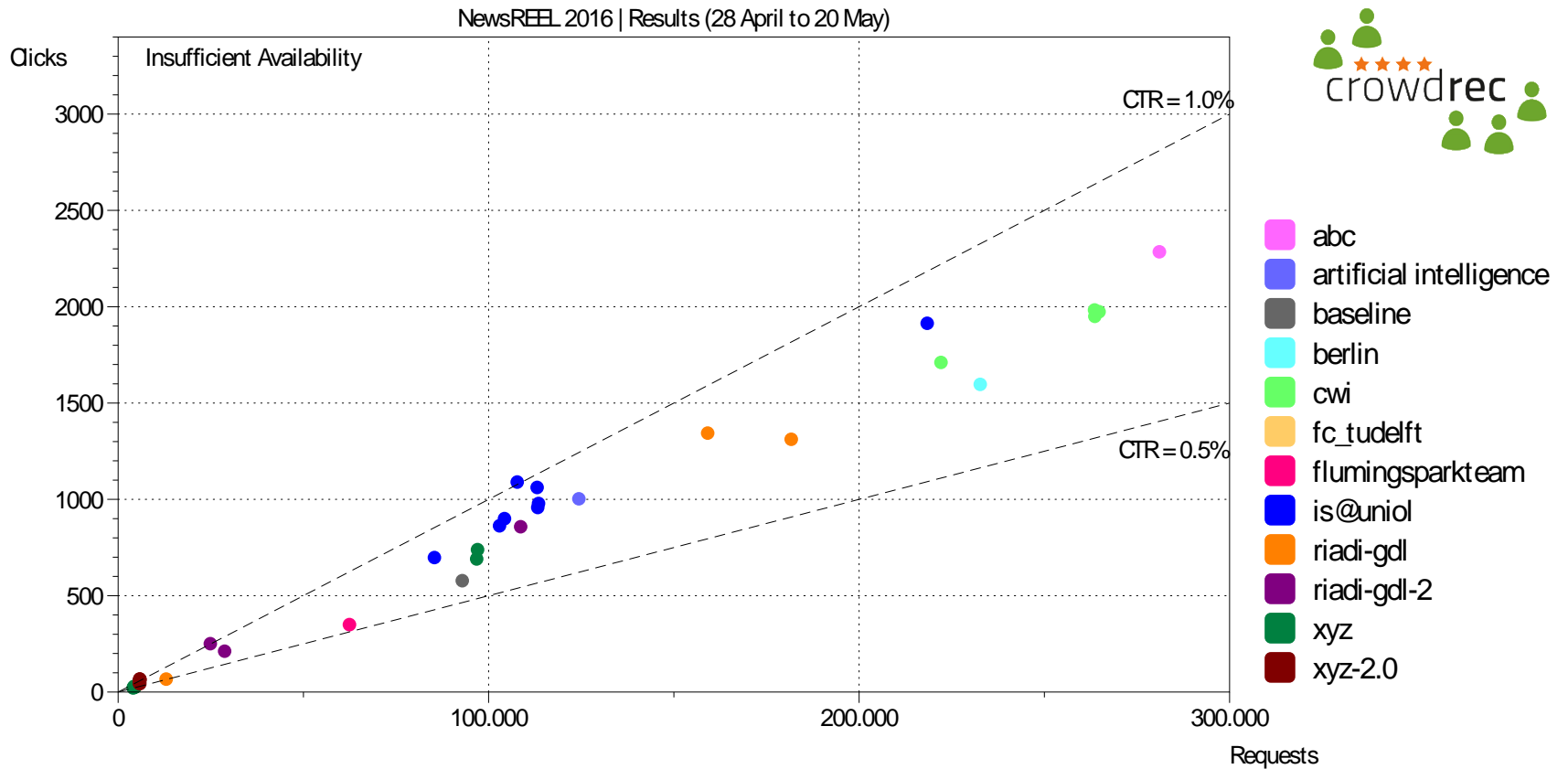
February 2016

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29					

March 2016

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

CTR (Task 1)



NewsREEL Session at CLEF 2016 in Evora

Winning algorithms were developed by students...



The image shows a framed award certificate for the CLEF NEWSREEL 2016 competition. The certificate is titled "CLEF NEWSREEL 2016" and "NEWS RECOMMENDATION EVALUATION LAB". It mentions the event was a "Campaign-style evaluation lab of CLEF 2016, Évora, Portugal, September 6th, 2016". The award is for "HIGHEST RECOMMENDATION PRECISION" and is presented to "PATRICK PROBST and ANDREAS LOMMATZSCH" for their work on "»Optimizing a Scalable News Recommender System«". The award is signed by Frank Hopfgartner on behalf of the organizers. The certificate features logos for crowdrec, CLEF ASSOCIATION, TU Berlin, plista, GRAVITY, TU Delft, moviri, and the University of Glasgow.

crowdrec **CLEF NEWSREEL 2016**
NEWS RECOMMENDATION EVALUATION LAB
Campaign-style evaluation lab of CLEF 2016, Évora, Portugal, September 6th, 2016

crowdrec **CLEF NEWSREEL 2016**
NEWS RECOMMENDATION EVALUATION LAB
Campaign-style evaluation lab of CLEF 2016, Évora, Portugal, September 6th, 2016

AWARD
HIGHEST RECOMMENDATION PRECISION

Is presented to
PATRICK PROBST and ANDREAS LOMMATZSCH
for the work described in the paper
»Optimizing a Scalable News Recommender System«

Frank Hopfgartner
On behalf of the organizers of NewsREEL 2016

CLEF ASSOCIATION
TU Berlin
plista
GRAVITY Research & Development
TU Delft Delft University of Technology
moviri
University of Glasgow

Overview

- Introduction
- Living Labs
- News Recommendation Evaluation Lab (NewsREEL) Scenario
- NewsREEL 2016
- **NewsREEL and You**

NewsREEL at CLEF 2017

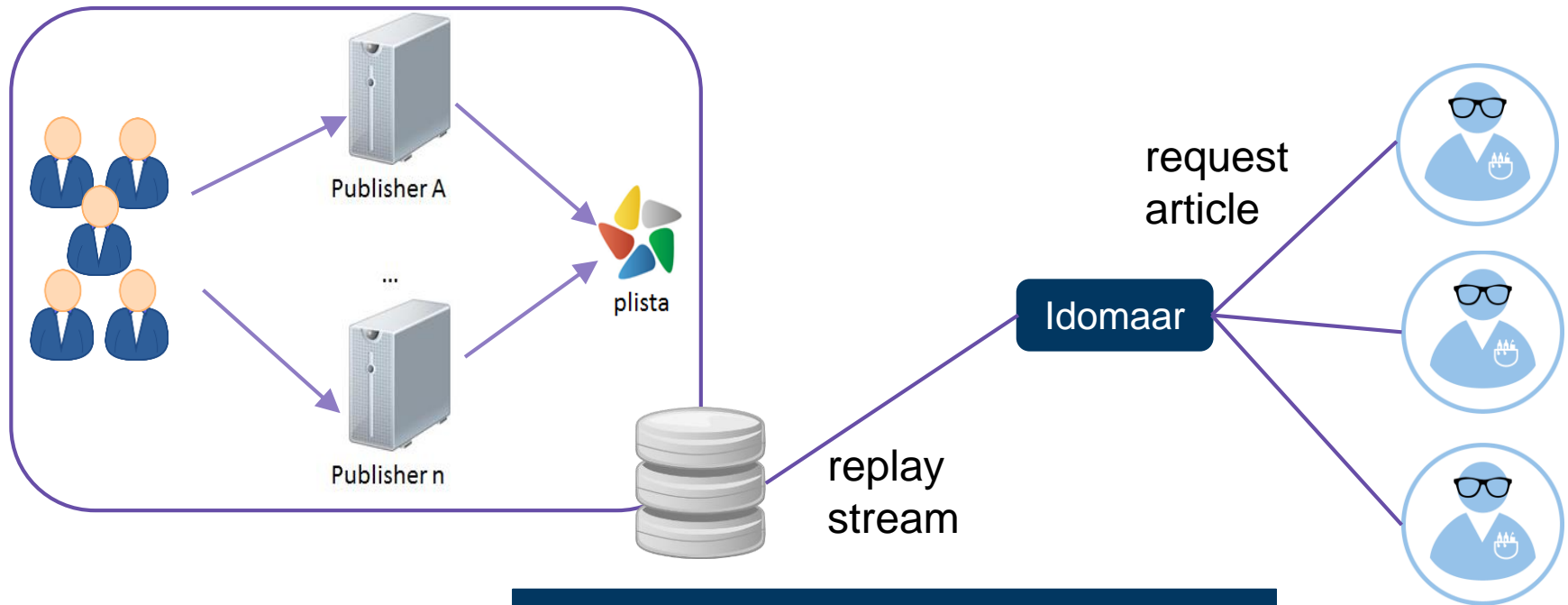
NewsREEL Live

- Provide **real-time** recommendations for visitors of news portals
- Benchmark own performance with other participants and baseline algorithms during **three pre-defined** evaluation windows
- Best algorithms determined in **final evaluation period**.

NewsREEL Replay

- Traffic and content updates of nine **news content provider** websites
- **Simulation** of data stream using **Idomaar framework**
- Participants have to **predict interactions** with data stream
- Metric: ratio of successful predictions by the total number of predictions

NewsREEL Replay



Eight news content providers
Two million notifications
58,000 item updates
168 million recommendation requests

NewsREEL for Learning & Teaching

“NewsREEL provided an excellent opportunity for my students to gain experience in using web services, processing streams, and developing information access algorithms.”

(Senior Academic)

Potentials of NewsREEL for Student Learning

Learning-by-doing

- Practical lab assignments as part of STEM curriculum to support self-referenced student learning and student engagement.

Limitations

- Limited complexity of current assignments results in “knowledge gap” between skillsets taught at HEIs and required by industry.

**Embed NewsREEL
in your teaching**

After participating in NewsREEL, students are able to...

The Potentials of Recommender Systems Challenges for Student Learning

Frank Hopfgartner, Andreas Lommatzsch, Benjamin Kille, Marthe Larson, Torben Brodt, Paolo Cremonesi, Alexandros Karatzoglou

CIML 2016, Barcelona, Spain

Teaching STEM courses

- Students of STEM courses need to gain skill sets that will allow them to thrive in industry. This includes technical skills but also skills for self-reflection and critical thinking.
- This requires the development of appropriate learning outcomes and the application of student-centered teaching methods that trigger students' intrinsic motivation to engage with these outcomes.
- Aim: Trigger deep learning by increased student engagement

Challenges

CLEF NewsREEL
<http://clef-newsreel.org/>

2016 RecSys Challenge
<http://2016.recsyschallenge.com/>



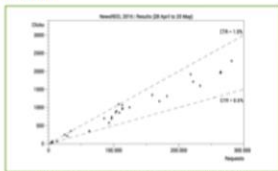
Enhance Student Engagement

- Theory:** Learning-by-doing activities can have a positive effect on student engagement.
- Common approach:** Introduce practical lab assignments as part of STEM curriculum, e.g., ask students of computer science courses to implement software to better understand techniques taught in the course.
- Problem:** Limited complexity of such lab assignments results in "knowledge gap".
- Suggestion:** Embrace potentials of campaign-style research challenges that are organized in conjunction with research conferences.

Examples – NewsREEL and RecSys'16 Challenge

- Current evaluation campaigns in the field of recommender systems
- NewsREEL:** Provide news recommendations for millions of visitors of selected online news portals. Benchmark algorithms in real-time via A/B testing or in a simulation-based setting.
- RecSys'16 Challenge:** Provide recommendations for users of a business and employment-oriented social networking service. Given a dataset, teams can train their algorithms and then submit solution.
- Both challenges attracted hundreds of researchers world-wide who gained an advanced technical skill set (e.g., using frameworks and libraries such as Apache Spark or Apache Flink), who at the same time familiarized themselves with academic practice of evaluation of different algorithms and techniques.

Benefits



Possibility to compare own performance against state-of-the-art (figure shows results of NewsREEL 2016 campaign)



Potential to add motivational layer of gamification, e.g., by letting students compete against classmates.



Results / Take Aways

- NewsREEL: Teachers predicting online streams.
- ACM RecSys Challenge: Teaches students to confront extreme sparsity using a variety of information sources.
- Success Stories: Plenty students successfully participated in 2016. Both challenges continue in 2017 in their next edition.

Further Reading

Benchmarking News Recommendations: The CLEF NewsREEL Use Case, SIGIR Forum, 49(2):129-136, 2016
F. Hopfgartner et al.



RecSys Challenge 2016: Job Recommendations, in RecSys'16, pp. 425-426, 2016
F. Abel et al.



Acknowledgement

This European Commission Project Contract Agreement no. 720944 funded the authors' part.



<http://www.clef-newsreel.org/> - <http://2016.recsyschallenge.com/>

- Demonstrate technical skills (e.g., how to handle data streams, recommendation techniques, use of Web protocols) needed to thrive in industry.
- Value the importance of critical evaluation and self reflection.
- Develop their own ideas and present them to an Academic audience.

Acknowledgements

People involved in NewsREEL (current and past)

CO-ORGANISERS

- Torben Brodt
- Andreas Lommatzsch
- Benjamin Kille
- Jonas Seiler
- Tobias Heintz
- Martha Larson
- Roberto Turrin

STEERING COMMITTEE

- Pablo Castells
- Paolo Cremonesi
- Arjen de Vries
- Michael Ekstrand
- Hideo Hoho
- Joemon M. Jose
- Noriko Kando
- Udo Kruschwitz
- Mounia Lalmas
- Jimmy Lin
- Vivien Petras
- Till Plumbaum
- Domonkos Tikk



Thank you

Questions?

Frank Hopfgartner
Frank.Hopfgartner@glasgow.ac.uk
@FTHopf
www.hopfgartner.co.uk