

Users2Story – On the Importance of Understanding Searchers' Information Needs

Udo Kruschwitz

School of Computer Science and Electronic Engineering
University of Essex
United Kingdom

udo@essex.ac.uk

26th March 2018

Why am I here?

Why am I here?

Here is a short answer:

Why am I here?

Here is a short answer: because I have been invited.

... but shouldn't I be there (NewsIR'18 Workshop)?

The University of Essex recently announced a new knowledge sharing deal which could have a major impact on the media monitoring market in the UK.

Under the KTP with technology company Signal Media, researchers at the university's Colchester campus will be involved in using artificial intelligence to convert large volumes of data from text files into more easily digestible formats.

The agreement aims to build on the success of a previous KTP between the university and Signal which helped the company secure investment of more than £5.8m and was named Best Partnership at the KTP Awards at the 2015 Innovate UK conference.

Professor Udo Kruschwitz, from the university's School of Computer Science and Electronic Engineering, will be the academic lead for the new partnership, which will see a postgraduate hired as a research associate and project manager for the KTP which is due to start in the third quarter of this year and to run for 18 months.

From left, Dr Miguel Martinez, co-founder and head of research at Signal Media, Professor Udo Kruschwitz from the School of Computer Science and Electronic Engineering at the University of Essex, and David Benigson, co-founder and chief executive of Signal Media, at the company's offices in Shoreditch, London.

Prof Kruschwitz said: "Our first KTP project has laid the groundwork for automatically processing, classifying and linking textual documents, and has evidently done so very successfully."

"We are now at a point where we will be moving from knowledge extraction to actual knowledge understanding and visualisation."

Picture: Adam Gasson

solved the containe through the next again by would maybe up by the French and de The French use on this en custo time the real s incomm amazi wrong

■ Bef make i real pr

■ Defi to the f the ans

■ Cons occurra to get te

M

So why am I here?

Topics of Text2Story

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining
- ▶ Narrative Summarization
- ▶ Storyline Visualization
- ▶ Temporal Aspects of Storylines
- ▶ Evaluation Methodologies for Narrative Extraction
- ▶ Big data applied to Narrative Extraction
- ▶ Resources and Dataset showcase
- ▶ Personalization and Recommendation
- ▶ User Profiling and User Behavior Modeling
- ▶ Credibility
- ▶ Fact Checking
- ▶ Bots Influence

Topics of Text2Story \cap My Expertise

Topics of Text2Story \cap My Expertise

- ▶ Event Identification

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining
- ▶ Narrative Summarization

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining
- ▶ Narrative Summarization
- ▶ Storyline Visualization

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining
- ▶ Narrative Summarization
- ▶ Storyline Visualization
- ▶ Temporal Aspects of Storylines

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining
- ▶ Narrative Summarization
- ▶ Storyline Visualization
- ▶ Temporal Aspects of Storylines
- ▶ Evaluation Methodologies for Narrative Extraction

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining
- ▶ Narrative Summarization
- ▶ Storyline Visualization
- ▶ Temporal Aspects of Storylines
- ▶ Evaluation Methodologies for Narrative Extraction
- ▶ Big data applied to Narrative Extraction

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining
- ▶ Narrative Summarization
- ▶ Storyline Visualization
- ▶ Temporal Aspects of Storylines
- ▶ Evaluation Methodologies for Narrative Extraction
- ▶ Big data applied to Narrative Extraction
- ▶ Resources and Dataset showcase

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining
- ▶ Narrative Summarization
- ▶ Storyline Visualization
- ▶ Temporal Aspects of Storylines
- ▶ Evaluation Methodologies for Narrative Extraction
- ▶ Big data applied to Narrative Extraction
- ▶ Resources and Dataset showcase
- ▶ **Personalization and Recommendation**
- ▶ **User Profiling and User Behavior Modeling**

Topics of Text2Story \cap My Expertise

- ▶ Event Identification
- ▶ Narrative Representation Language
- ▶ Sentiment and Opinion Detection
- ▶ Argumentation Mining
- ▶ Narrative Summarization
- ▶ Storyline Visualization
- ▶ Temporal Aspects of Storylines
- ▶ Evaluation Methodologies for Narrative Extraction
- ▶ Big data applied to Narrative Extraction
- ▶ Resources and Dataset showcase
- ▶ **Personalization and Recommendation**
- ▶ **User Profiling and User Behavior Modeling**
- ▶ Credibility
- ▶ Fact Checking
- ▶ Bots Influence

Conclusions

Conclusions

... well

Conclusions

... well ... conclusions so far ...

Conclusions

Conclusions

- ▶ User profiling fits the scope of the workshop

Conclusions

- ▶ User profiling fits the scope of the workshop
- ▶ It is one of many interesting aspects to be looked at

Conclusions

- ▶ User profiling fits the scope of the workshop
- ▶ It is one of many interesting aspects to be looked at
- ▶ Profiling on the Web?

Conclusions

- ▶ User profiling fits the scope of the workshop
- ▶ It is one of many interesting aspects to be looked at
- ▶ Profiling on the Web? I leave that to Google *et al.*

Conclusions

- ▶ User profiling fits the scope of the workshop
- ▶ It is one of many interesting aspects to be looked at
- ▶ Profiling on the Web? I leave that to Google *et al.*
(and Cambridge Analytica / NSA / GCHQ ...)

Conclusions

- ▶ User profiling fits the scope of the workshop
- ▶ It is one of many interesting aspects to be looked at
- ▶ Profiling on the Web? I leave that to Google *et al.*
(and Cambridge Analytica / NSA / GCHQ ...)
- ▶ Profiling in smaller collections?

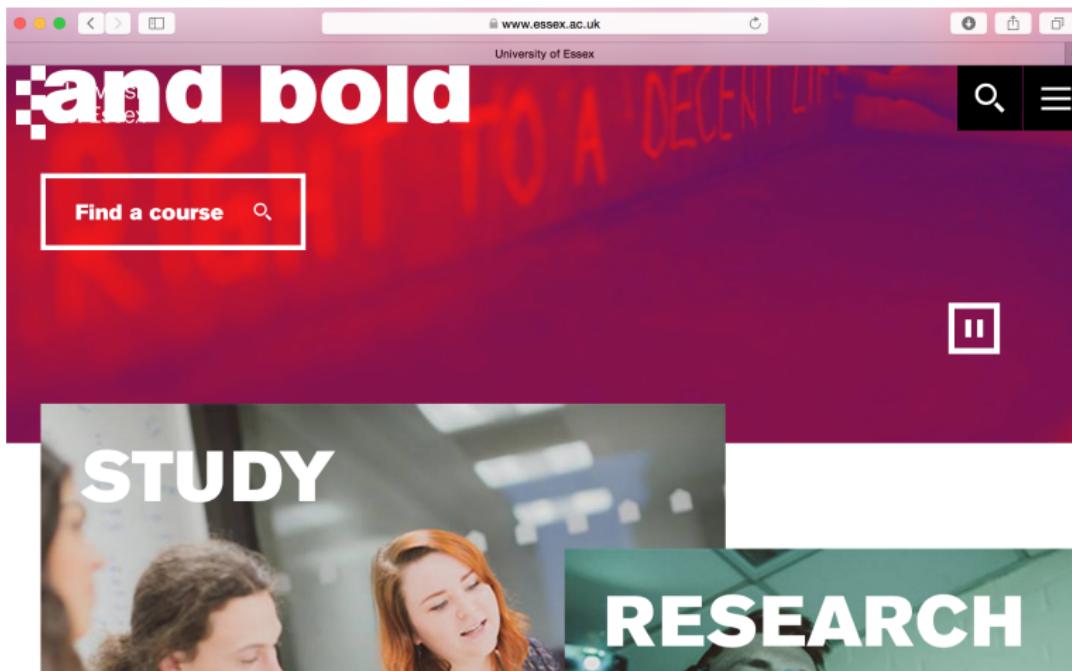
Conclusions

- ▶ User profiling fits the scope of the workshop
- ▶ It is one of many interesting aspects to be looked at
- ▶ Profiling on the Web? I leave that to Google *et al.*
(and Cambridge Analytica / NSA / GCHQ ...)
- ▶ Profiling in smaller collections? That's my use case (*intranet search, enterprise search, site search, professional search*)

Conclusions

- ▶ User profiling fits the scope of the workshop
- ▶ It is one of many interesting aspects to be looked at
- ▶ Profiling on the Web? I leave that to Google *et al.*
(and Cambridge Analytica / NSA / GCHQ ...)
- ▶ Profiling in smaller collections? That's my use case (*intranet search, enterprise search, site search, professional search*)
- ▶ So let's look at some motivating examples ...

Motivation



How do I find anything?

Searching!

Searching!

The screenshot shows a web browser window for the University of Essex Course Search. The URL in the address bar is <https://www.essex.ac.uk/course-search?query=AI>. The page features the University of Essex logo and a search bar. The main content area displays a message: "COURSE RESULTS FOR: 'AI'" followed by "No results found". Below this message, a sub-message states: "We're sorry but we can't find any results that match your search term." At the bottom of the page are two buttons: "Find a course" and a search bar placeholder "Search again".

COURSE RESULTS FOR:
'AI'

No results found

We're sorry but we can't find any results that match your search term.

Find a course Search again

Browse courses by subject

Keep Searching

Keep Searching

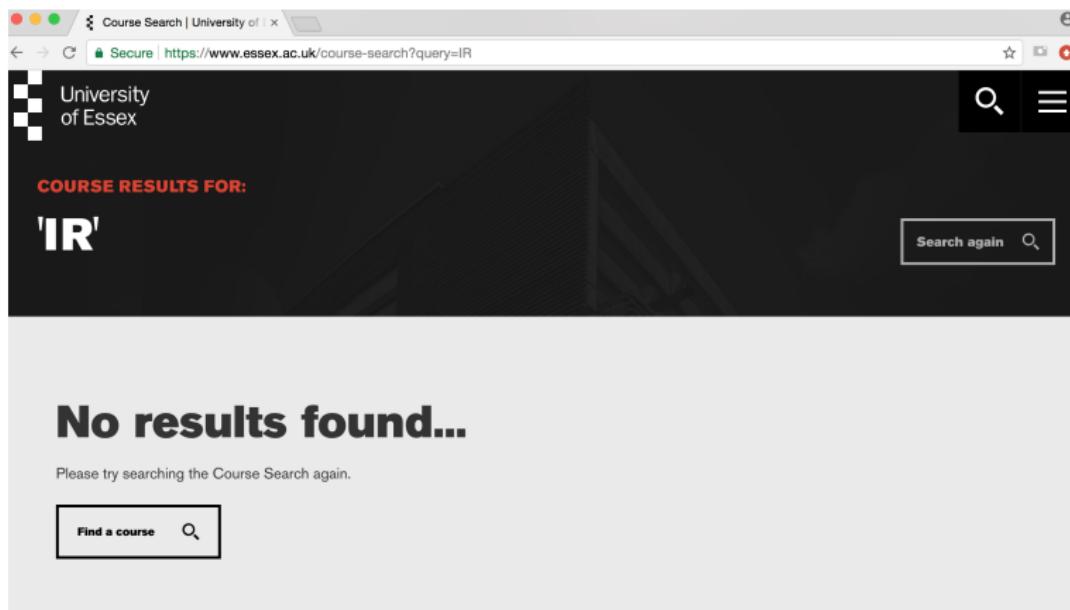
The screenshot shows a web browser window with the following details:

- Title Bar:** Course Search | University of Essex
- Address Bar:** Secure | https://www.essex.ac.uk/course-search?query=natural+language+processing
- University Logo:** University of Essex
- Search Bar:** A magnifying glass icon and a menu icon.
- Section Header:** COURSE RESULTS FOR:
- Search Query:** 'natural language processing'
- Search Again Button:** A button labeled "Search again" with a magnifying glass icon.
- Main Content:** A large bold text "No results found..."
- Text Below:** Please try searching the Course Search again.
- Bottom Buttons:** "Find a course" button and a search bar with a magnifying glass icon.

... and Searching ...

The screenshot shows a web browser window for the University of Essex Course Search. The URL is https://www.essex.ac.uk/course-search?query=nlp. The page displays a search interface with the University of Essex logo and a search bar containing the query 'nlp'. Below the search bar, a red banner reads 'COURSE RESULTS FOR: 'nlp''. A large, bold message 'No results found...' is centered on the page, followed by the text 'Please try searching the Course Search again.' At the bottom left is a 'Find a course' button, and at the bottom right is a search bar with a magnifying glass icon.

... and Searching ...



... found something!

... found something!

The screenshot shows a web browser window with the following details:

- Title Bar:** Search | University of Essex
- Address Bar:** Secure | https://www.essex.ac.uk/search?query=IR
- University of Essex Logo:** A black and white checkered logo next to the text "University of Essex".
- Search Bar:** A magnifying glass icon and a three-line menu icon.
- Search Results Header:** SITE RESULTS FOR: 'IR'
- Search Again Button:** A button labeled "Search again" with a magnifying glass icon.
- Left Sidebar (Narrow your results):** Categories: All (215), Webpages (191), Research Repository (22), Academic Staff (2).
- Main Content Area:** 1-10 of 215 results. The title "All Results (215)" is displayed in large bold letters.
- Result 1 (Webpage):** WEBPAGE, The University of Essex Language and Computation Group - Language and Co...
Description: of language and computation — including Computational Linguistics (CL), Natural Language Processing (NLP), Information Retrieval (IR) and Formal Semantics — then do feel free to contact us.
Link: http://lac.essex.ac.uk/
- Result 2 (Research Repository):** RESEARCH REPOSITORY, Impulse radio Ultrawideband D2D-based localization for ultra-dense 5G netwo...
Description: (Summary text cut off)

... found even more ...

www.essex.ac.uk/search?collection=use-meta&query=ir&start_rank=81

Search | University of Essex

WEBPAGE
g]kfnfsf] ;z:q áGádf afnaflnsflj?4 a9\bfl] lx+f hgj/L @)% o'4sf] ...

IR, N? 2. 5)#!, ? 4, ? 4, ,(=; 3 1)# >. ((>, 8 B S#S, 1#H..((>, 8 B S#S, 1#, 7 \$H:
https://www1.essex.ac.uk/armedcon/story_id/000250.pdf

WEBPAGE
SaloneBP0711final.PDF

https://www1.essex.ac.uk/armedcon/story_id/The%20Jury%20is%20still%20out.pdf

in Guinea, Liberia or within Sierra Leone are returning home to rebuild their lives; civilians abducted from their villages during rebel attacks are being reunited with their families; the

https://www1.essex.ac.uk/armedcon/story_id/The%20Jury%20is%20still%20out.pdf

WEBPAGE
Working at height: A brief guide

ng s. ys te. m s. eg. a ir. ba gs.,
<https://www1.essex.ac.uk/health-safety/activities/documents/working-height-brief-guide.pdf>

... unfortunately not what I was looking for

The screenshot shows a web browser window with three search results listed vertically. Each result is enclosed in a light gray box and includes a small thumbnail icon on the left.

- WEBPAGE chap49.dvi**
Version of 31.8.09. Chapter 49. Further topics. I conclude the volume with notes on six almost unconnected special topics. In 491 I look at equidis- tributed sequences and the ideal Z of sets with asymptotic density zero. I give the principal
<https://www1.essex.ac.uk/mathematics/people/fremlin/chap49.pdf>
- WEBPAGE chap25.dvi**
Version of 31.5.03. Chapter 25. Product Measures. I come now to another chapter on 'pure' measure theory, discussing a fundamental construction – or, as you may prefer to consider it, two constructions, since the problems involved in forming
<https://www1.essex.ac.uk/mathematics/people/fremlin/chap25.pdf>
- WEBPAGE chap46.dvi**
Version of 26.8.13. Chapter 46. Pointwise compact sets of measurable functions. This chapter collects results inspired by functional analysis. 461 and 466 look directly at measures on linear topological spaces. The primary applications
<https://www1.essex.ac.uk/mathematics/people/fremlin/chap46.pdf>

Observation

- ▶ Information (in intranets / Web sites / other collections) can be difficult to find

Core Problem: Enterprise/Site/Intranet \neq Web¹

¹A lot more on this in (Kruschwitz & Hull, 2017)

Core Problem: Enterprise/Site/Intranet \neq Web¹

- ▶ Smaller scale, fewer users

¹A lot more on this in (Kruschwitz & Hull, 2017)

Core Problem: Enterprise/Site/Intranet \neq Web¹

- ▶ Smaller scale, fewer users
- ▶ Lack of link structure

¹A lot more on this in (Kruschwitz & Hull, 2017)

Core Problem: Enterprise/Site/Intranet \neq Web¹

- ▶ Smaller scale, fewer users
- ▶ Lack of link structure
- ▶ Documents often created for simple dissemination

¹A lot more on this in (Kruschwitz & Hull, 2017)

Core Problem: Enterprise/Site/Intranet \neq Web¹

- ▶ Smaller scale, fewer users
- ▶ Lack of link structure
- ▶ Documents often created for simple dissemination
- ▶ Often just a single relevant document (or none at all)

¹A lot more on this in (Kruschwitz & Hull, 2017)

Core Problem: Enterprise/Site/Intranet \neq Web¹

- ▶ Smaller scale, fewer users
- ▶ Lack of link structure
- ▶ Documents often created for simple dissemination
- ▶ Often just a single relevant document (or none at all)
- ▶ Little redundancy

¹A lot more on this in (Kruschwitz & Hull, 2017)

Core Problem: Enterprise/Site/Intranet \neq Web¹

- ▶ Smaller scale, fewer users
- ▶ Lack of link structure
- ▶ Documents often created for simple dissemination
- ▶ Often just a single relevant document (or none at all)
- ▶ Little redundancy
- ▶ Documents in silos with different access restrictions

¹A lot more on this in (Kruschwitz & Hull, 2017)

Core Problem: Enterprise/Site/Intranet \neq Web¹

- ▶ Smaller scale, fewer users
- ▶ Lack of link structure
- ▶ Documents often created for simple dissemination
- ▶ Often just a single relevant document (or none at all)
- ▶ Little redundancy
- ▶ Documents in silos with different access restrictions
- ▶ ...

¹A lot more on this in (Kruschwitz & Hull, 2017)

Addressing the Problem

Addressing the Problem

- ▶ Observation: many searchers with similar information needs

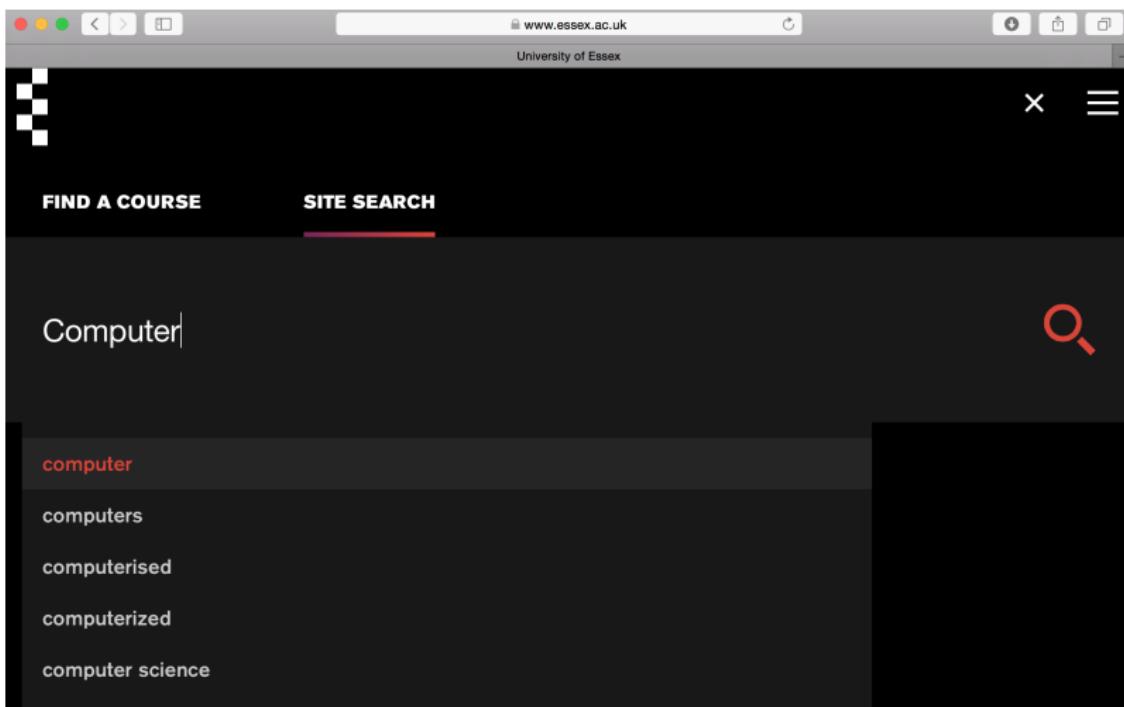
Addressing the Problem

- ▶ Observation: many searchers with similar information needs
- ▶ How about tapping into the *wisdom of the crowd* (exploiting the fact that some searchers are actually successful)?

Addressing the Problem

- ▶ Observation: many searchers with similar information needs
- ▶ How about tapping into the *wisdom of the crowd* (exploiting the fact that some searchers are actually successful)?
- ▶ Utilise this (somehow ‘understanding’ the underlying information needs) to assist others with similar requests

Could be used for Search ...



... as well as Navigation

The screenshot shows the University of Essex website. At the top, there is a navigation bar with links: Study with us, Life at Essex, International, Departments, and Research. The 'Departments' link is highlighted. Below the navigation bar, the word 'Departments' is prominently displayed. A list of academic departments is provided, including Biological Sciences, Computational Finance and Economic Agents, Computer Science and Electronic Engineering (which is highlighted with a mouse cursor), East 15 Acting School, Economics, Edge Hotel School, Essex Business School, Government, and Health and Human Sciences. To the right of this list are three more groups of departments: History, Human Rights, and Interdisciplinary Studies; Literature, Film, and Theatre Studies; and Mathematical Sciences. A callout box with a blue border and white text is overlaid on the page, pointing to the 'Computer Science and Electronic Engineering' link. The text in the callout box reads: 'We focus on providing you with the skills and experience that you need in the graduate job market. Our courses are designed to enable you to enter a wide variety of careers and our graduates have gone on to work for several well-known employers.' At the bottom of the page, there are links for Law and University of Essex Online.

- > Biological Sciences
- > Computational Finance and Economic Agents
- > **Computer Science and Electronic Engineering**
- > East 15 Acting School
- > Economics
- > Edge Hotel School
- > Essex Business School
- > Government
- > Health and Human Sciences

We focus on providing you with the skills and experience that you need in the graduate job market. Our courses are designed to enable you to enter a wide variety of careers and our graduates have gone on to work for several well-known employers.

- > History
- > Human Rights
- > Interdisciplinary Studies
- > Literature, Film, and Theatre Studies
- > Mathematical Sciences

- > Law
- > University of Essex Online

Building a Profile

- ▶ Utilise the search trails of users
- ▶ Turn them into a useful knowledge structure (automatically)
- ▶ No individual *user profiles* but profiles reflecting an entire *cohort* of users
- ▶ Query/navigation logs are a good starting point

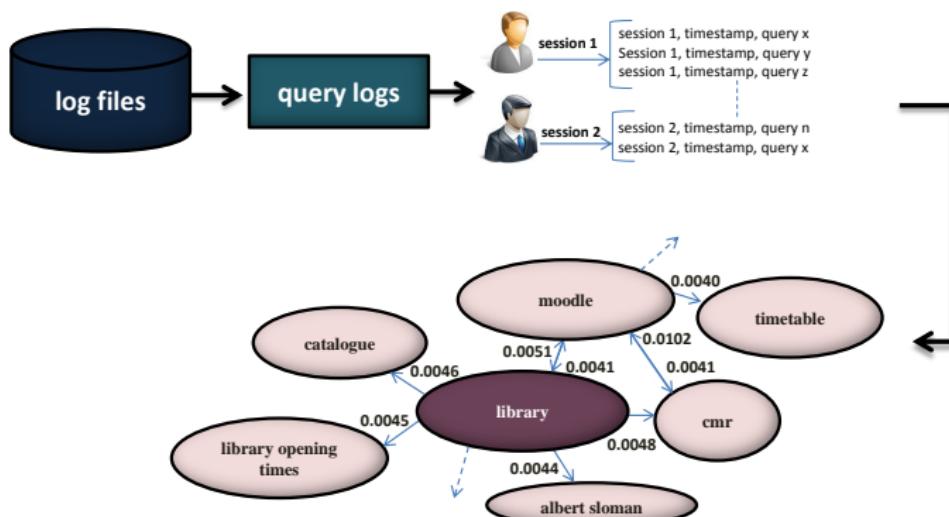
How exactly do we mine the Log Data?

Log Data Example

```
...
1657406 xxx 02f4e527278fbe909aedb97cea27f8d9 \
    xxx wed nov 17 11:41:43 gmt 2010 xxx 0 xxx 0 xxx 0 \
    xxx librarary xxx librarary xxx librarary
...
1657409 xxx 02f4e527278fbe909aedb97cea27f8d9 \
    xxx wed nov 17 11:41:58 gmt 2010 xxx 0 xxx 0 xxx 0 \
    xxx albert salmon xxx albert salmon xxx albert salmon
...
1657412 xxx 02f4e527278fbe909aedb97cea27f8d9 \
    xxx wed nov 17 11:42:23 gmt 2010 xxx 0 xxx 0 xxx 0 \
    xxx albert salmon library xxx albert salmon library \
    xxx albert salmon library
...
1657415 xxx 02f4e527278fbe909aedb97cea27f8d9 \
    xxx wed nov 17 11:42:37 gmt 2010 xxx 0 xxx 0 xxx 0 \
    xxx albert sloman library xxx albert sloman library \
    xxx albert sloman library
...

```

From Logs to Profile



From Logs to Profile II

From Logs to Profile II

- ▶ Queries submitted by users

From Logs to Profile II

- ▶ Queries submitted by users
- ▶ Identify sessions

From Logs to Profile II

- ▶ Queries submitted by users
- ▶ Identify sessions
- ▶ Associate related queries

From Logs to Profile II

- ▶ Queries submitted by users
- ▶ Identify sessions
- ▶ Associate related queries (many possible ways of doing so)

From Logs to Profile II

- ▶ Queries submitted by users
- ▶ Identify sessions
- ▶ Associate related queries (many possible ways of doing so)
- ▶ Result is a query association graph

From Logs to Profile II

- ▶ Queries submitted by users
- ▶ Identify sessions
- ▶ Associate related queries (many possible ways of doing so)
- ▶ Result is a query association graph (of some sort)

From Logs to Profile II

- ▶ Queries submitted by users
- ▶ Identify sessions
- ▶ Associate related queries (many possible ways of doing so)
- ▶ Result is a query association graph (of some sort)
- ▶ Work reported here uses a three-year *query* log collected on a university search engine comprising about 1.6 million queries

From Logs to Profile III

Variety of models, including:

From Logs to Profile III

Variety of models, including:

- ▶ Maximum Likelihood Estimates (MLE)

From Logs to Profile III

Variety of models, including:

- ▶ Maximum Likelihood Estimates (MLE)
- ▶ Formal Concept Analysis (FCA)

From Logs to Profile III

Variety of models, including:

- ▶ Maximum Likelihood Estimates (MLE)
- ▶ Formal Concept Analysis (FCA)
- ▶ Ant Colony Optimization analogy (ACO)

From Logs to Profile III

Variety of models, including:

- ▶ Maximum Likelihood Estimates (MLE)
- ▶ Formal Concept Analysis (FCA)
- ▶ Ant Colony Optimization analogy (ACO)
- ▶ Enhanced Query Flow Graph (QFG)

From Logs to Profile III

Variety of models, including:

- ▶ Maximum Likelihood Estimates (MLE)
- ▶ Formal Concept Analysis (FCA)
- ▶ Ant Colony Optimization analogy (ACO)
- ▶ Enhanced Query Flow Graph (QFG)
- ▶ Hybrid Approach: Documents + Query Logs

From Logs to Profile III

Variety of models, including:

- ▶ Maximum Likelihood Estimates (MLE)
- ▶ Formal Concept Analysis (FCA)
- ▶ Ant Colony Optimization analogy (ACO)
- ▶ Enhanced Query Flow Graph (QFG)
- ▶ Hybrid Approach: Documents + Query Logs

... no time to look at any of these approaches ...

Profile-Based Summarisation for Web Site Navigation²

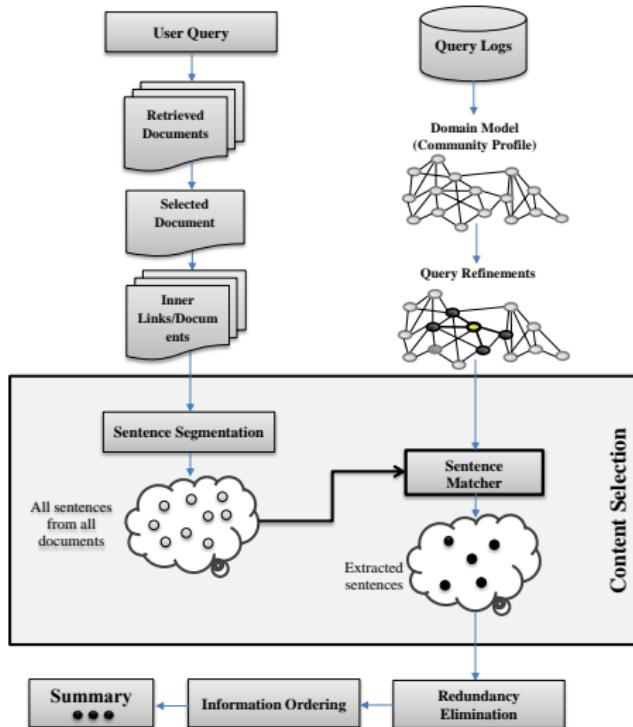
²Details in (Alhindi *et al.*, 2015)

Profile-Based Summarisation for Web Site Navigation²

- ▶ Motivation: assist a user in navigating a Web site (using hover text and automatic summarisation)
- ▶ Applying ACO to summarise documents as you are browsing
- ▶ Single document vs. multi-document summaries

²Details in (Alhindi *et al.*, 2015)

Profile-based Summarisation



Profile-Based Summary for Navigation

The screenshot shows the University of Essex website's navigation bar and a detailed view of the 'Departments' section. The navigation bar includes links for 'Study with us', 'Life at Essex', 'International', 'Departments', and 'Research'. The 'Departments' link is highlighted with a yellow box. Below this, the 'Departments' section lists various academic units. A callout box highlights the 'Computer Science and Electronic Engineering' link, which is underlined and has a mouse cursor pointing to it. The callout box contains a descriptive text about the department's focus on providing skills for the graduate job market.

University of Essex STUD

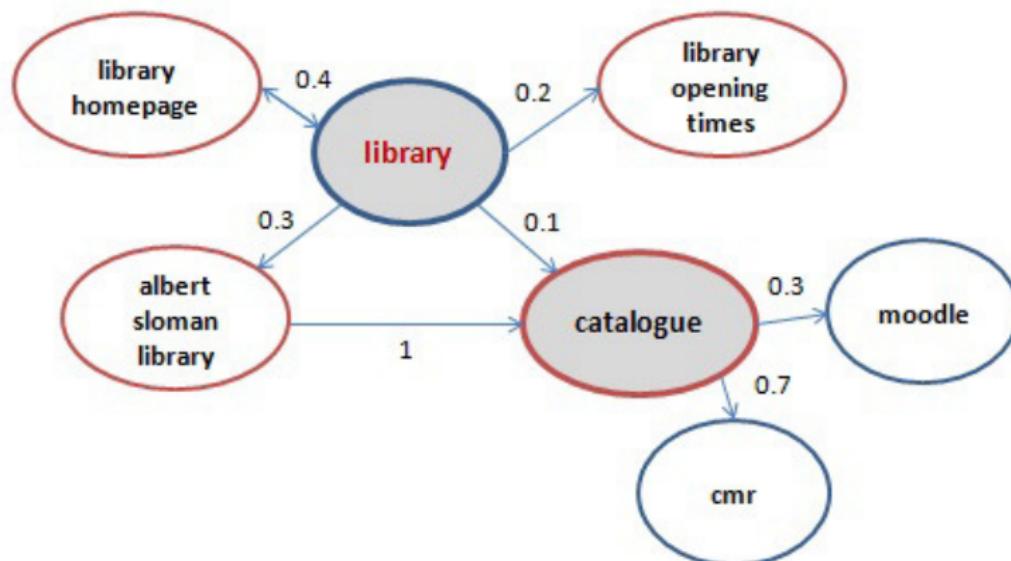
Study with us Life at Essex International Departments Research

Departments

- › Biological Sciences
- › Computational Finance and Economic Agents
- › **Computer Science and Electronic Engineering**
- › East 15 Acting School
- › Economics
- › Edge Hotel School
- › Essex Business School
- › Government
- › Health and Human Sciences
- › History
- › Human Rights
- › Interdisciplinary Studies
- › Law
- › Literature, Film, and Theatre Studies
- › Mathematical Sciences
- › University of Essex Online

We focus on providing you with the skills and experience that you need in the graduate job market. Our courses are designed to enable you to enter a wide variety of careers and our graduates have gone on to work for several well-known employers.

Reminder: Partial Profile



Assessing the Acquired Profile(s)³

- ▶ Sampled frequent and less frequent queries
- ▶ User study to assess quality of derived term suggestions using a variety of log-based and document-based methods

³(Kruschwitz et al., 2013) in JASIST

Assessing the Acquired Profile(s)³

- ▶ Sampled frequent and less frequent queries
- ▶ User study to assess quality of derived term suggestions using a variety of log-based and document-based methods (e.g. extracting suggestions from snippets of best matches)

³(Kruschwitz et al., 2013) in JASIST

Assessing the Acquired Profile(s)³

- ▶ Sampled frequent and less frequent queries
- ▶ User study to assess quality of derived term suggestions using a variety of log-based and document-based methods (e.g. extracting suggestions from snippets of best matches, association rules approach)

³(Kruschwitz et al., 2013) in JASIST

Assessing the Acquired Profile(s)³

- ▶ Sampled frequent and less frequent queries
- ▶ User study to assess quality of derived term suggestions using a variety of log-based and document-based methods (e.g. extracting suggestions from snippets of best matches, association rules approach, query flow graphs)

³(Kruschwitz et al., 2013) in JASIST

Assessing the Acquired Profile(s)³

- ▶ Sampled frequent and less frequent queries
- ▶ User study to assess quality of derived term suggestions using a variety of log-based and document-based methods (e.g. extracting suggestions from snippets of best matches, association rules approach, query flow graphs, maximum likelihood estimation)

³(Kruschwitz et al., 2013) in JASIST

Assessing the Acquired Profile(s)³

- ▶ Sampled frequent and less frequent queries
- ▶ User study to assess quality of derived term suggestions using a variety of log-based and document-based methods (e.g. extracting suggestions from snippets of best matches, association rules approach, query flow graphs, maximum likelihood estimation...)

³(Kruschwitz et al., 2013) in JASIST

Assessing the Acquired Profile(s)³

- ▶ Sampled frequent and less frequent queries
- ▶ User study to assess quality of derived term suggestions using a variety of log-based and document-based methods (e.g. extracting suggestions from snippets of best matches, association rules approach, query flow graphs, maximum likelihood estimation...)
- ▶ Maximum Likelihood approach very accurate (but sparse)

³(Kruschwitz et al., 2013) in JASIST

Assessing the Acquired Profile(s)³

- ▶ Sampled frequent and less frequent queries
- ▶ User study to assess quality of derived term suggestions using a variety of log-based and document-based methods (e.g. extracting suggestions from snippets of best matches, association rules approach, query flow graphs, maximum likelihood estimation...)
- ▶ Maximum Likelihood approach very accurate (but sparse)
- ▶ Query Flow Graphs better coverage and consistent

³(Kruschwitz et al., 2013) in JASIST

Assessing the Acquired Profile(s)³

- ▶ Sampled frequent and less frequent queries
- ▶ User study to assess quality of derived term suggestions using a variety of log-based and document-based methods (e.g. extracting suggestions from snippets of best matches, association rules approach, query flow graphs, maximum likelihood estimation...)
- ▶ Maximum Likelihood approach very accurate (but sparse)
- ▶ Query Flow Graphs better coverage and consistent
- ▶ Session-based approach seems ok, but more fine-grained session identification is better

³(Kruschwitz et al., 2013) in JASIST

Search Application: Longitudinal Study at BT⁴

- ▶ Applying profile for query suggestion in a realistic call centre setting

⁴(Albakour et al., 2013)

Search Application: Longitudinal Study at BT⁴

- ▶ Applying profile for query suggestion in a realistic call centre setting
- ▶ BT engineers both in the field and the call centre

⁴(Albakour et al., 2013)

Search Application: Longitudinal Study at BT⁴

- ▶ Applying profile for query suggestion in a realistic call centre setting
- ▶ BT engineers both in the field and the call centre
- ▶ Search engine indexing different information silos
- ▶ A/B testing applied to MLE, AR and ACO

⁴(Albakour et al., 2013)

Search Application: Longitudinal Study at BT⁴

- ▶ Applying profile for query suggestion in a realistic call centre setting
- ▶ BT engineers both in the field and the call centre
- ▶ Search engine indexing different information silos
- ▶ A/B testing applied to MLE, AR and ACO
- ▶ Finding: low uptake , but ...

⁴(Albakour et al., 2013)

Search Application: Longitudinal Study at BT⁴

- ▶ Applying profile for query suggestion in a realistic call centre setting
- ▶ BT engineers both in the field and the call centre
- ▶ Search engine indexing different information silos
- ▶ A/B testing applied to MLE, AR and ACO
- ▶ Finding: low uptake , but ...
- ▶ Higher uptake of ACO suggestions than the alternatives
- ▶ Statistically significant increase after training phase

⁴(Albakour et al., 2013)

Search Application: Longitudinal Study at BT⁴

- ▶ Applying profile for query suggestion in a realistic call centre setting
 - ▶ BT engineers both in the field and the call centre
 - ▶ Search engine indexing different information silos
 - ▶ A/B testing applied to MLE, AR and ACO
 - ▶ Finding: low uptake , but ...
 - ▶ Higher uptake of ACO suggestions than the alternatives
 - ▶ Statistically significant increase after training phase
- ... more details in the book chapter

⁴(Albakour et al., 2013)

Task-Based Evaluation of Navigation

- ▶ Compare three systems
- ▶ Sample task:

You are going to be a new postgraduate student at the University of Essex. You need to locate a page with useful information about tuition fees and possible funding offered by the university.

Task-Based Evaluation of Navigation

- ▶ Compare three systems
- ▶ Sample task:

You are going to be a new postgraduate student at the University of Essex. You need to locate a page with useful information about tuition fees and possible funding offered by the university.

- ▶ TREC-style task-based evaluation (TREC Interactive Track)
- ▶ 18 subjects each conducting 6 tasks
- ▶ Series of questionnaires and additional metrics (time taken, number of turns)

Task-Based Evaluation of Navigation

- ▶ Compare three systems
- ▶ Sample task:

You are going to be a new postgraduate student at the University of Essex. You need to locate a page with useful information about tuition fees and possible funding offered by the university.

- ▶ TREC-style task-based evaluation (TREC Interactive Track)
- ▶ 18 subjects each conducting 6 tasks
- ▶ Series of questionnaires and additional metrics (time taken, number of turns)
- ▶ All the details in the TOIS paper ...

Task-Based Evaluation: Setup

The screenshot shows the University of Essex website. At the top, there's a navigation bar with links: Study with us, Life at Essex, International, Departments (which is highlighted in blue), and Research. Below this, a secondary navigation bar has links: Departments, Online, and Book. A tooltip is overlaid on the 'Departments' link in the second bar. The tooltip contains the following text:

Our school has six computer laboratories exclusively for computer science and electronic engineering students. Computing and electronics are major industries both in the UK and across the globe and demand for graduates of these disciplines is outstripping supply.

Figure: System B.

The screenshot shows the University of Essex website. The layout is similar to System B, with a top navigation bar and a secondary navigation bar below it. A tooltip is overlaid on the 'Departments' link in the secondary bar. The tooltip contains the following text:

We have an outstanding reputation for teaching and high-quality research in both computer science and electronic engineering. In the national student satisfaction survey 2013, we were ranked first in the UK for computer science and fourth for electronic engineering.

Figure: System C.

Evaluation Results

- ▶ Profiles of user cohorts can assist future users with similar requests

Evaluation Results

- ▶ Profiles of user cohorts can assist future users with similar requests
- ▶ Profile-based summarisation significantly reduces number of steps and time to find documents

Evaluation Results

- ▶ Profiles of user cohorts can assist future users with similar requests
- ▶ Profile-based summarisation significantly reduces number of steps and time to find documents
- ▶ Multi-document summaries are marginally better than single-document summaries

Conclusions

- ▶ Important differences between intranet/site/... and internet

Conclusions

- ▶ Important differences between intranet/site/... and internet
- ▶ Users2Story: capturing information needs of searchers in a profile (to assist future searchers with similar requests)

Conclusions

- ▶ Important differences between intranet/site/... and internet
- ▶ Users2Story: capturing information needs of searchers in a profile (to assist future searchers with similar requests)
- ▶ Log files are a useful resource to build profiles representing user cohorts

Conclusions

- ▶ Important differences between intranet/site/... and internet
- ▶ Users2Story: capturing information needs of searchers in a profile (to assist future searchers with similar requests)
- ▶ Log files are a useful resource to build profiles representing user cohorts
- ▶ Profiles can be used for search as well as browsing/navigation

Conclusions

- ▶ Important differences between intranet/site/... and internet
- ▶ Users2Story: capturing information needs of searchers in a profile (to assist future searchers with similar requests)
- ▶ Log files are a useful resource to build profiles representing user cohorts
- ▶ Profiles can be used for search as well as browsing/navigation
- ▶ Adaptive models (e.g. ant colony optimization) learn over time and do improve

Conclusions

- ▶ Important differences between intranet/site/... and internet
- ▶ Users2Story: capturing information needs of searchers in a profile (to assist future searchers with similar requests)
- ▶ Log files are a useful resource to build profiles representing user cohorts
- ▶ Profiles can be used for search as well as browsing/navigation
- ▶ Adaptive models (e.g. ant colony optimization) learn over time and do improve
- ▶ Usual problems: noise, sparsity, ...

Acknowledgements

- ▶ Azhar Alhindi
- ▶ Dyaa Albakour (who is next door in the NewsIR'18 workshop)
- ▶ Chris Fox
- ▶ ...

References (ACO in BT Prototype)

- ▶ M-D. Albakour, G. Ducatel, and U. Kruschwitz. *The Role of Search for Field Force Knowledge Management*. In Transforming Field and Service Operations, pages 117–132. Springer. 2013.

References (Navigation)

- ▶ A. Alhindi, U. Kruschwitz, C. Fox & M-D. Albakour. *Profile-based Summarisation for Web Site Navigation*. ACM Transactions on Information Systems (TOIS), Special Issue on Contextual Search and Recommendation. 33(1): 4:1-4:39, March 2015.
- ▶ S. Z. Saad & U. Kruschwitz. *Applying Web Usage Mining for Adaptive Intranet Navigation*. In Proceedings of the 2nd Information Retrieval Facility Conference, volume 6653 of LNCS, Springer, 2011.
- ▶ S. Z. Saad & U. Kruschwitz. *Exploiting Click Logs for Adaptive Intranet Navigation*. In Proceedings of ECIR 2013, Moscow.

References (Query Suggestion Evaluation)

- ▶ U. Kruschwitz, D. Lungley, M-D. Albakour and D. Song
Deriving Query Suggestions for Site Search. Journal of the American Society for Information Science and Technology (JASIST), 64(10):1975–1994, October 2013.

In other News: Out now ...

Foundations and Trends® in Information Retrieval
Vol. 11, No. 1 (2017) 1–142
© 2017 U. Kruschwitz and C. Hull
DOI: 10.1561/1500000053



Searching the Enterprise

Udo Kruschwitz
University of Essex, UK
udo@essex.ac.uk

Charlie Hull
Flax, UK
charlie@flax.co.uk

See you at SIGIR?

The screenshot shows the homepage of the ProfS2018 website. At the top, there is a navigation bar with the logo "ProfS2018" and five green buttons labeled "OVERVIEW", "CFP", "PROGRAM", "ORGANISATION", and "CONTACT". Below the navigation bar, the main title "First International Workshop on Professional Search" is displayed in large white font. Underneath the title, the text "Co-located with SIGIR 2018, Ann Arbor, Michigan, USA -- 12 July 2018" is shown in white. The bottom section of the page has a dark background and contains the heading "Overview" in white, followed by a paragraph of text.

ProfS2018

OVERVIEW CFP PROGRAM ORGANISATION CONTACT

First International Workshop on Professional Search

Co-located with SIGIR 2018, Ann Arbor, Michigan, USA --
12 July 2018

Overview

Professional search in specific domains has been addressed in IR research over the last decades. Although each domain (e.g. legal, medical, academic, governmental) has its own idiosyncrasies, professional search tasks have specific requirements in common that are different from requirements of generic web search engines. These requirements follow directly from the context and needs of professional searchers. Searchers in different domains often exhibit particular