Context Aware Predictive Analytics:

Motivation, Potential, Challenges

Mykola Pechenizkiy

31 October 2011

University of Bournemouth, England

http://www.win.tue.nl/~mpechen/projects/CAPA

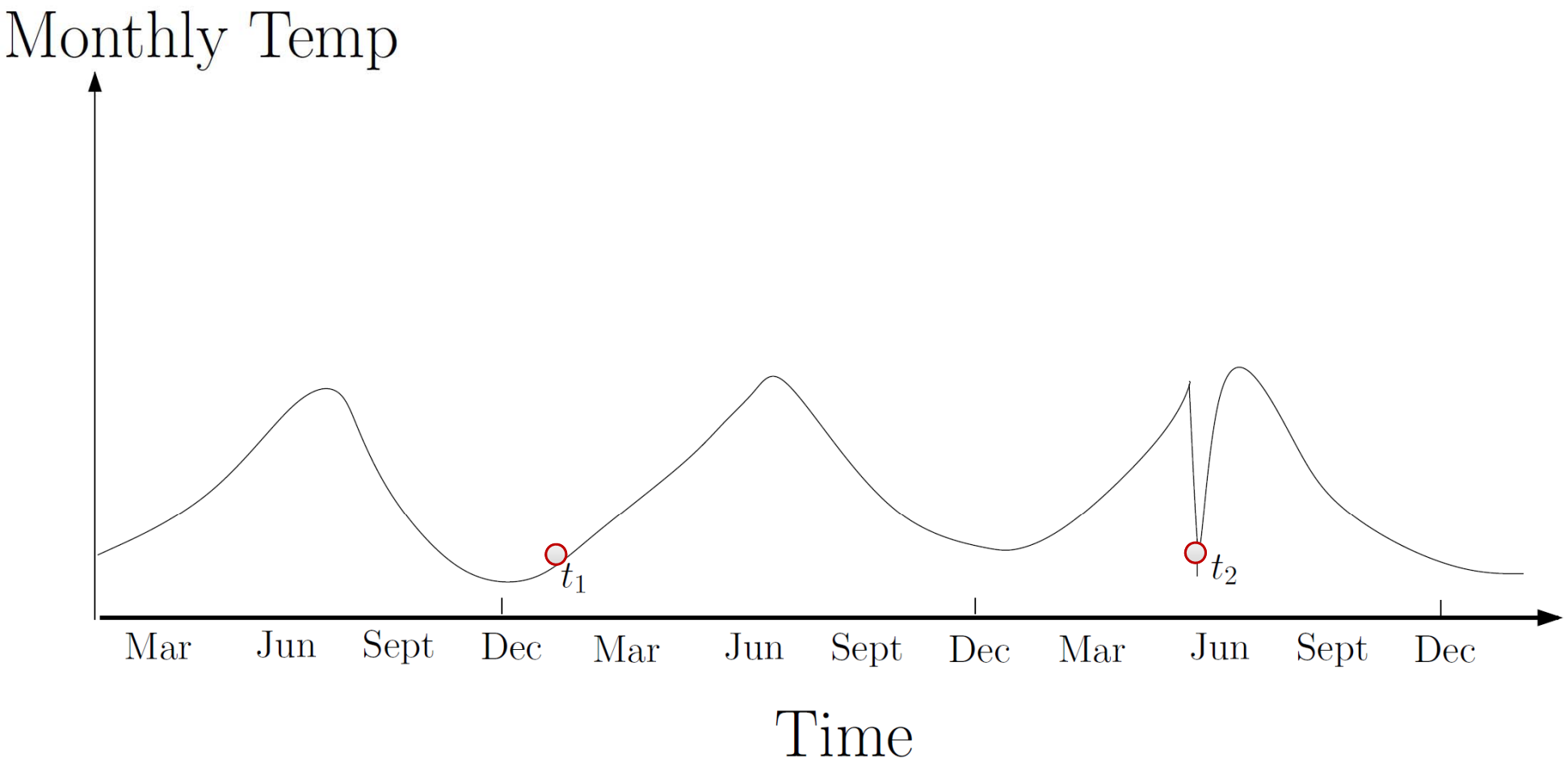
Outline

- What is context? In supervised learning tasks?
- Examples of why context is important in different applications
- Context-Aware Predictive Analytics framework
  - Motivation for multilevel model design & (re)training
  - Research questions and current directions
- CAPA in Web Analytics:
  - Advertisement, recommender systems, and alike;
  - Two major tasks: demand prediction (popularity) & relevance prediction (personal recommendations);
  - Importance of different data sources integration

What is Context?

- Definition of context depends on the context …
  - Many (formal & vague) definitions in different fields
- Context as any *additional* information that
  - enhances the understanding of the instance of interest,
  - and in particular helps us to classify this instance
- We do not need _the_ formal definition of context in this talk;
  - instead we'll consider several intuitive example

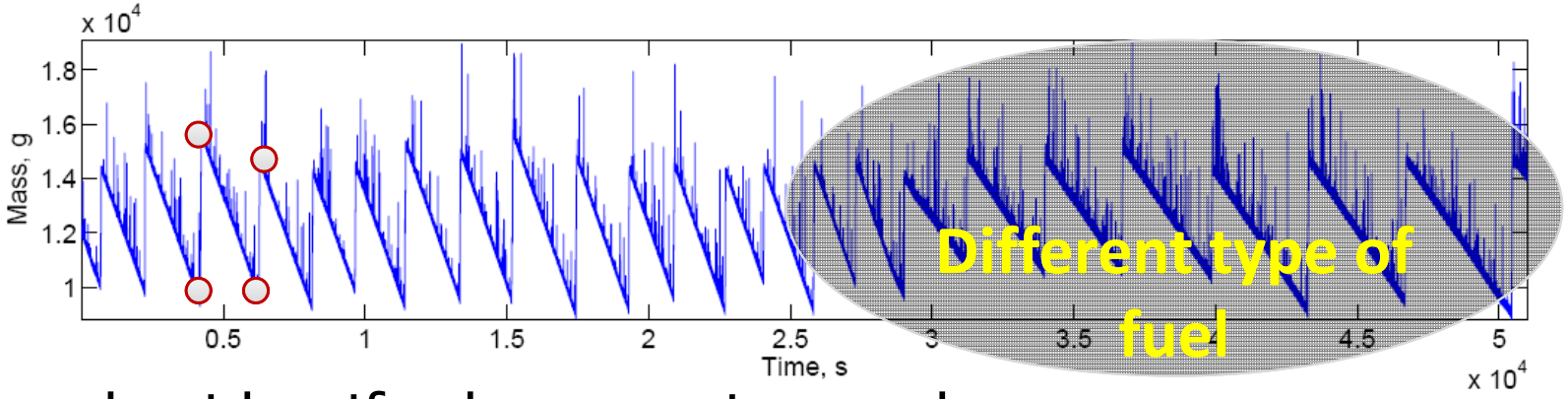
Context in Outlier Detection



If we have seasonality, simple thresholding does not work; need to adjust w.r.t. the season

Context in Change Detection

Online Mass Flow Prediction in CFB boiler:

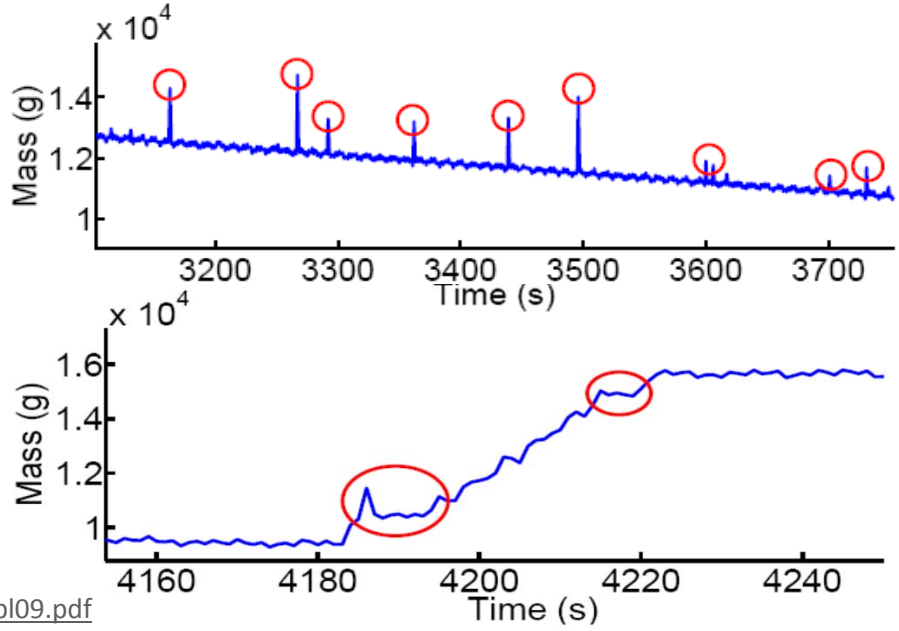


Subtasks: identify change points and distinguish them from outliers

Both are difficult due to:

- asymmetric nature of the outliers

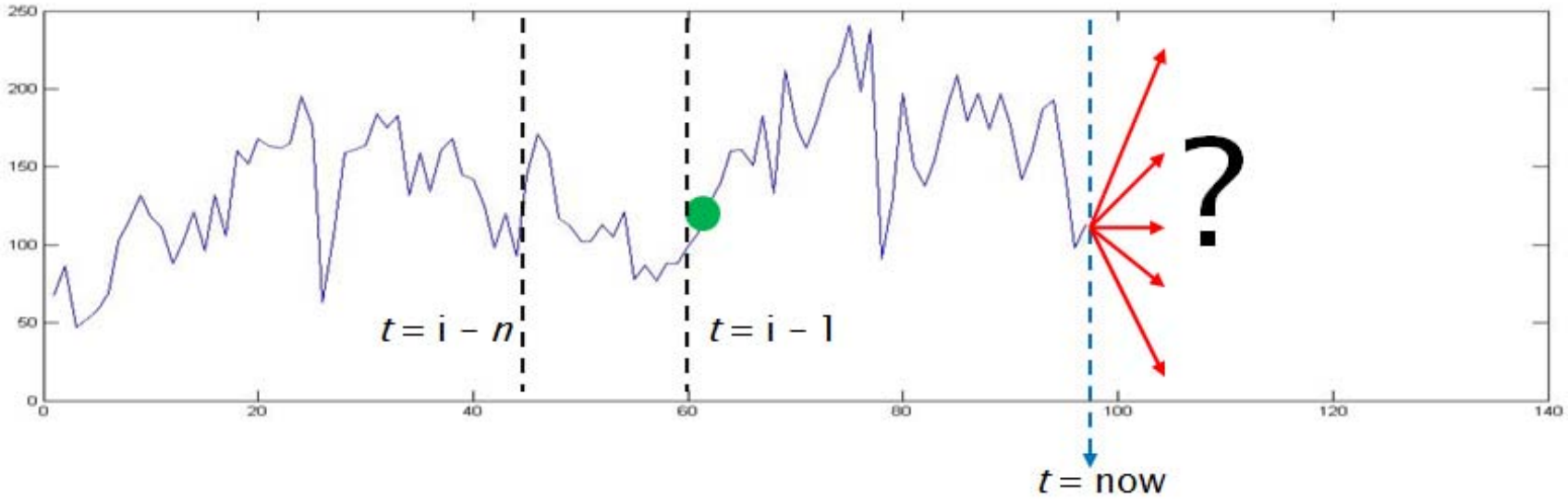
- short consumption periods within feeding stages



http://www.win.tue.nl/~mpechen/publications/pubs/PechenizkiySIGKDDExpl09.pdf

Context in Food Sales Prediction

Contextual features characterize/categorize behavior of a product



t	History	Temp	Holiday	Promo
1				
.				
i	{y(i-n) .. y(i-1)}			
.				
n				

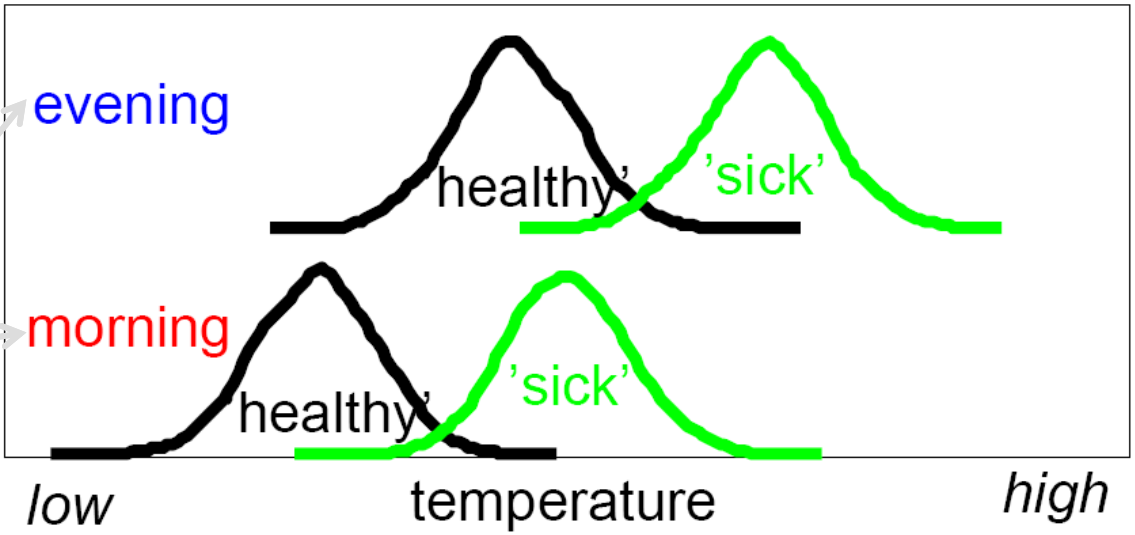
Predict label of each new instance

http://dx.doi.org/10.1016/j.eswa.2011.07.078

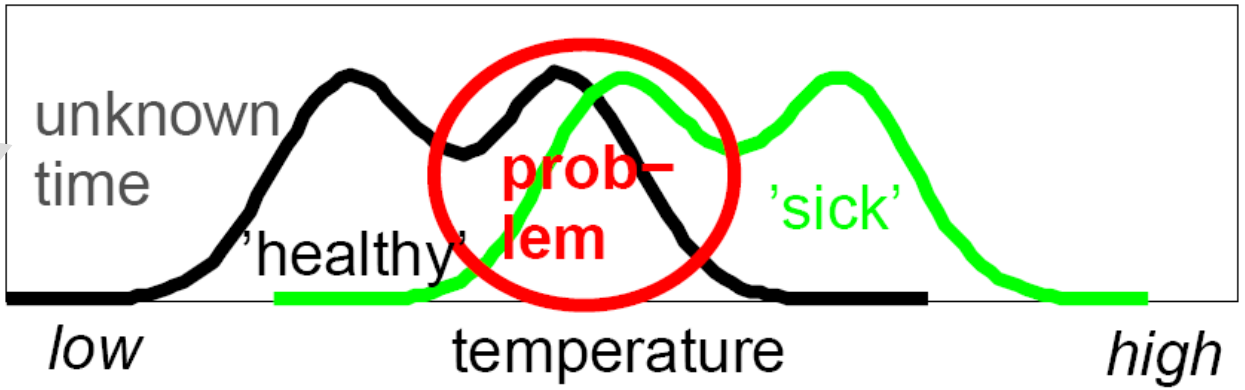
Context in Classification

Not predictive alone but a subset of features with the contextual attribute(s) becomes (much) more predictive

Time of the day context



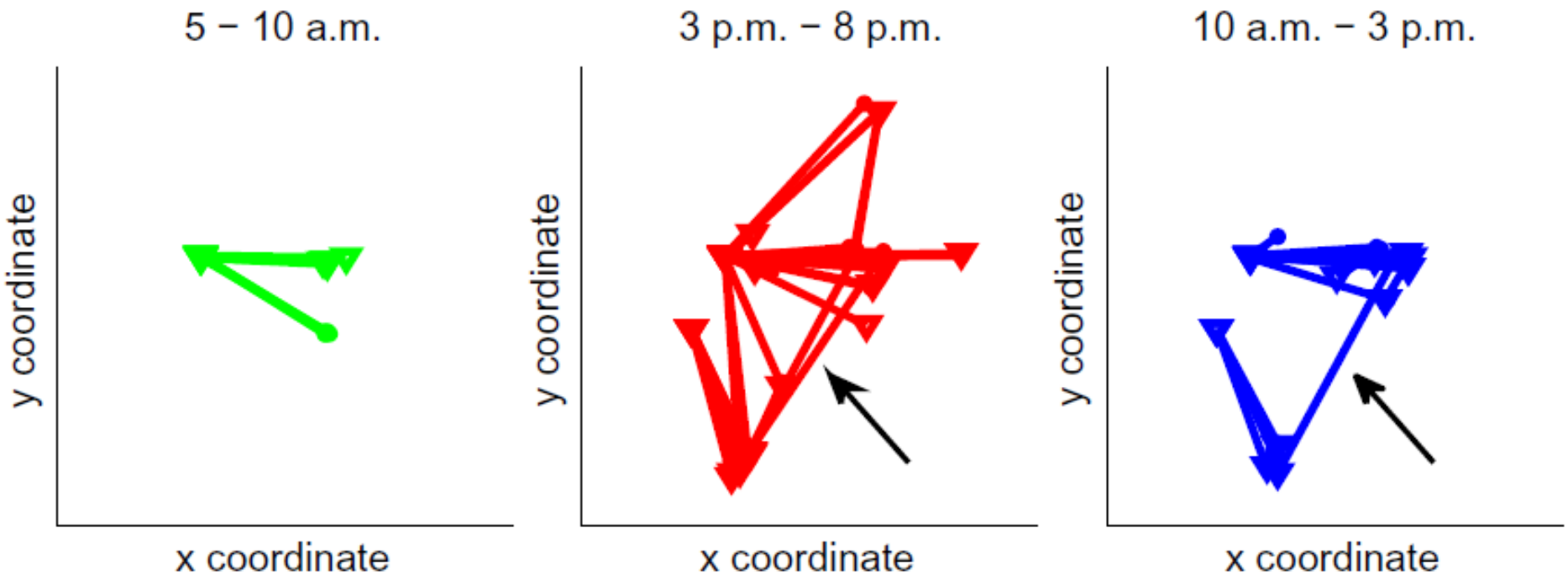
no context



https://sites.google.com/site/zliobaite/Zliobaite_PAKDD11_CR.pdf

Context-aware Driving Route Recognition

Baseline approach: LCSS matching on the database of existing reference routes



Time of day context

http://www.springerlink.com/content/1k540l8741324137/

Context in Face Recognition

Variety of poses, lighting, facial expression, partial occlusions

- Finding a model or a single way to learn a model that recognizes face/identifies a person under such diverse settings is difficult

Context may not effect/identify an object, but can help to choose a better preprocessing technique, classifier etc

- Just adding background features to the representation space is not enough, i.e. multi-level prediction is needed

But e.g. “has glasses” can be a contextual attribute that e.g. suggests a model not to rely on the position of eyes, but can be also a predictive attribute /prior of belonging to a particular class

Context in Word Sense Disambiguation

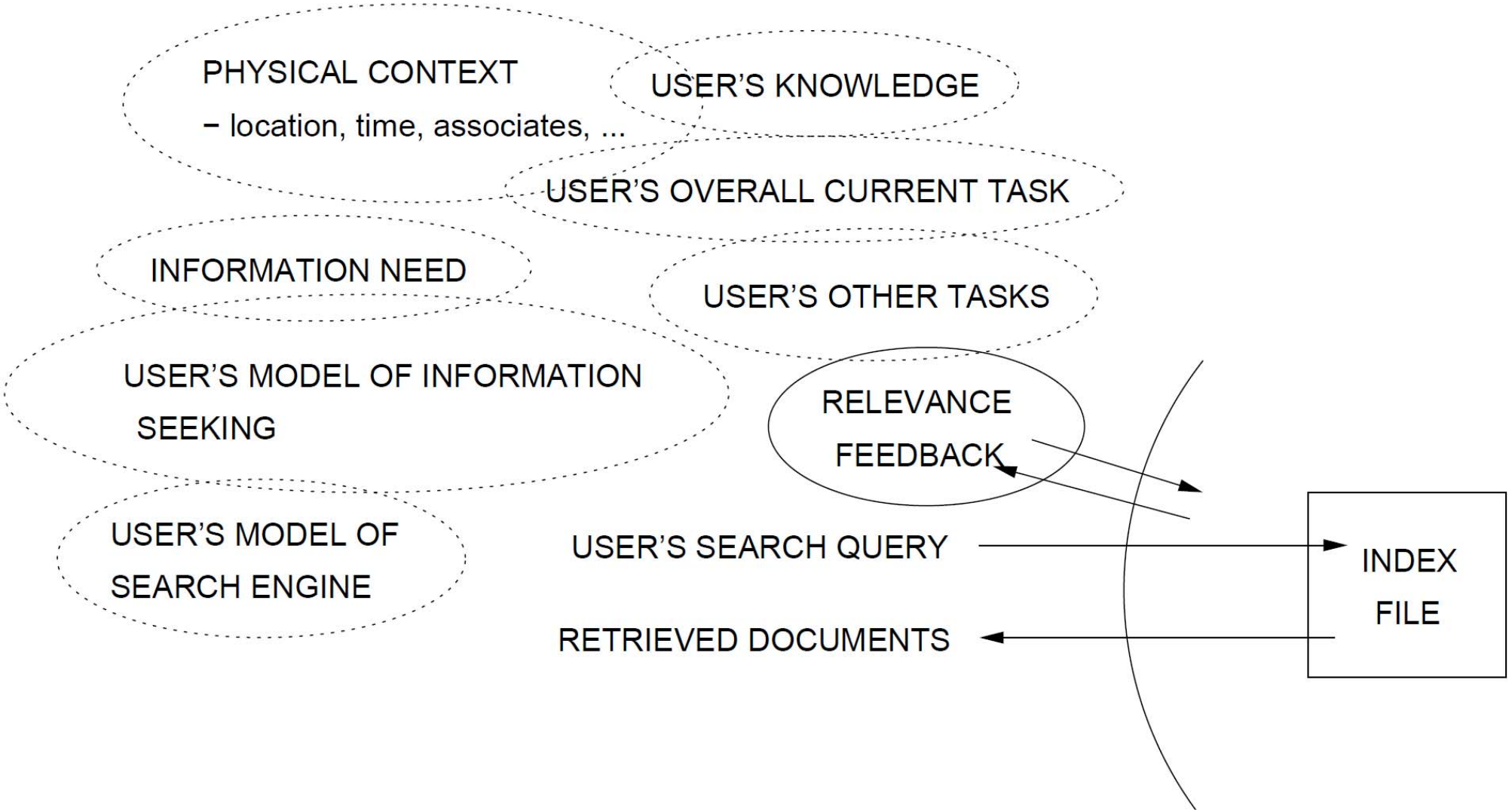
I went fishing for some sea bass.
The bass line of the song is too weak.

- WSD - the process of identifying which sense of a word (i.e. meaning) is used in a sentence, when the word has multiple meanings (polysemy)
- (Un/Semi-)Supervised methods are based on the assumption that the context can provide enough evidence on its own to disambiguate words
  - Which are predictive features?
  - Which are contextual?

Context in IR & RecSys

- User Context
  - Preferences, usage history, profiles
- Document/Product Context
  - Meta-data, content features
- Task Context
  - Current activity, location etc.
- Social Context
  - Leveraging the social graph

This can be refined further:



Jones & Brown, 2004 The Role of Context in Information Retrieval

Mobile IR

- Profile, Location, Time,
- Activity, Agenda, Service,
- Preferences, Situation,
- Environment, Social Context

Context is certainly important …

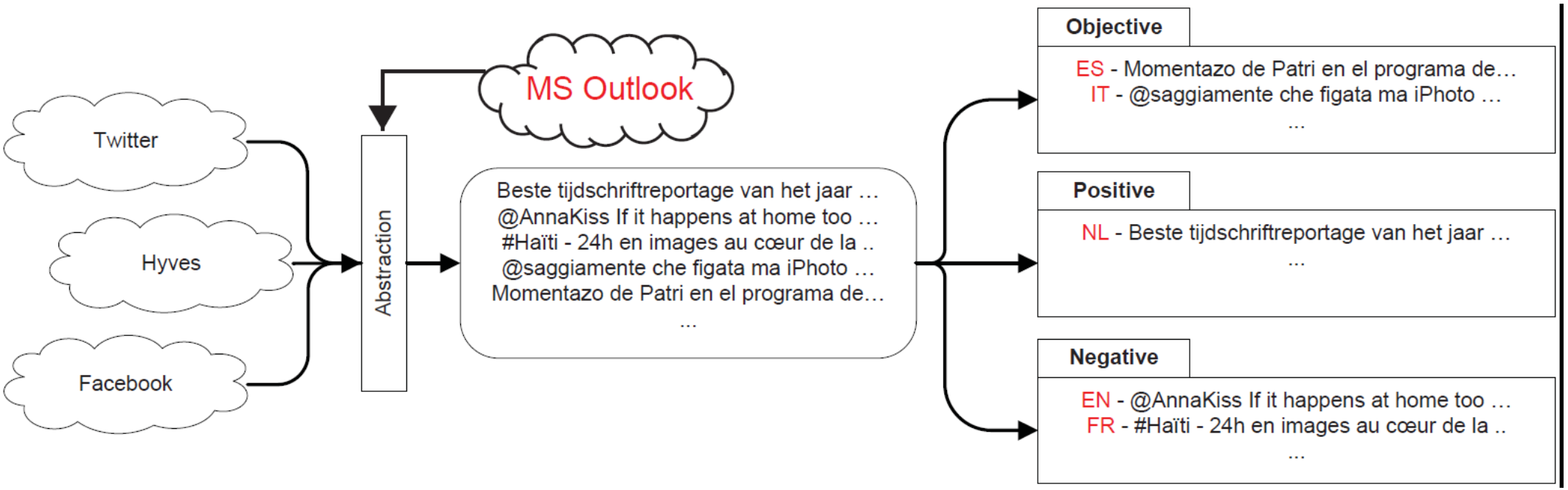
… but it is not obvious what elements of context are useful, when and why!

Learning vs. Engineering CAPA

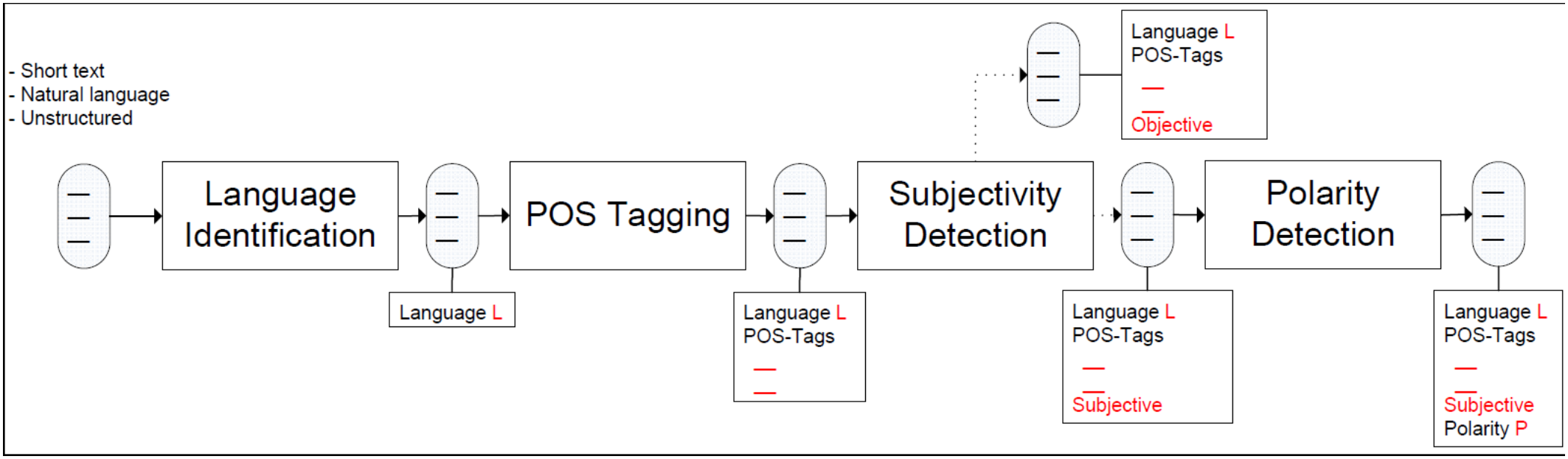
- Examples of engineering of context-aware/multi-level classification systems
  - Multi-lingual sentiment classification, emotion recognition
  - Relevant content identification in a web-page

- What can/should we design and what should we aim to induce from the data?

Sentiment Classification



Four Step Sentiment Classification



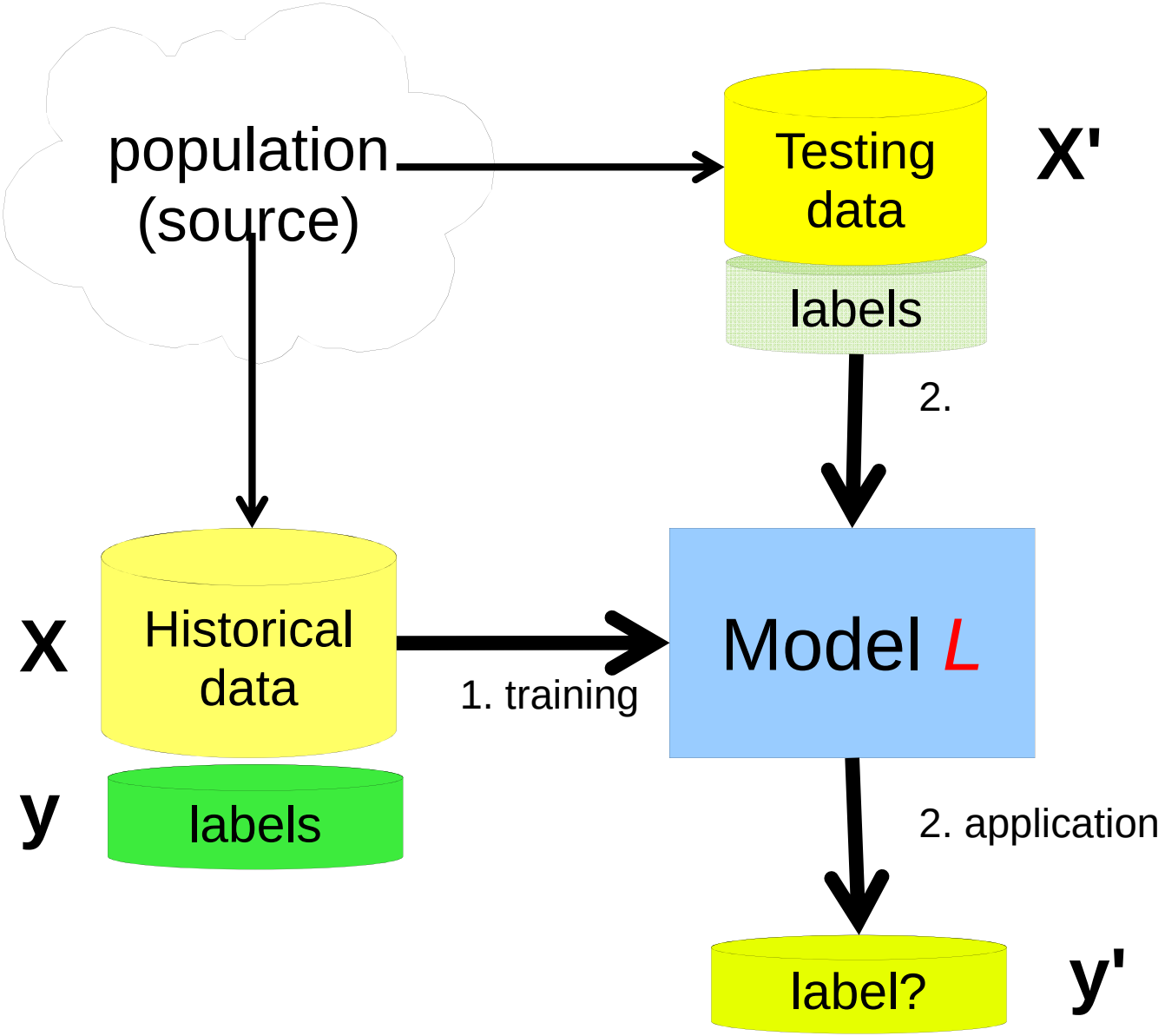
- Multi-level decision making: seemingly more complicated process in fact simplifies the learning task by explicitly pointing out to the algorithm what to do
- Similar ideas are practiced a lot in many domains where we want to introduce the knowledge of the domain into the mining process, data fusion, information fusion

Identifying Content Worth Indexing

1st: Find web page segments

2nd: Decide whether a segment contains relevant info

Supervised Learning



Training:

$y = L(X)$

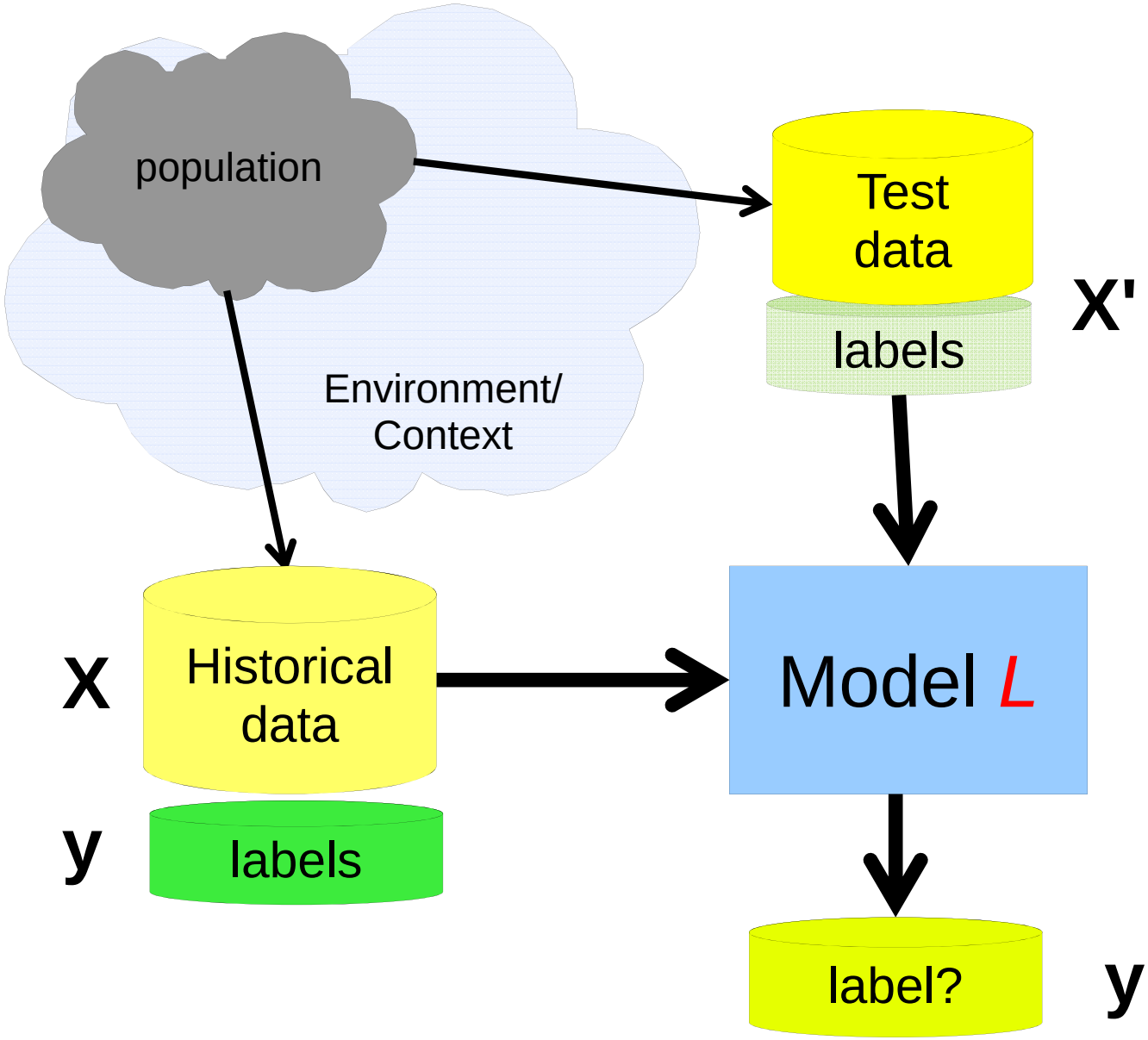
Application:

use *L*

for an unseen data

$y' = L(X')$

Sup.Learning with Context-Awareness



Training:

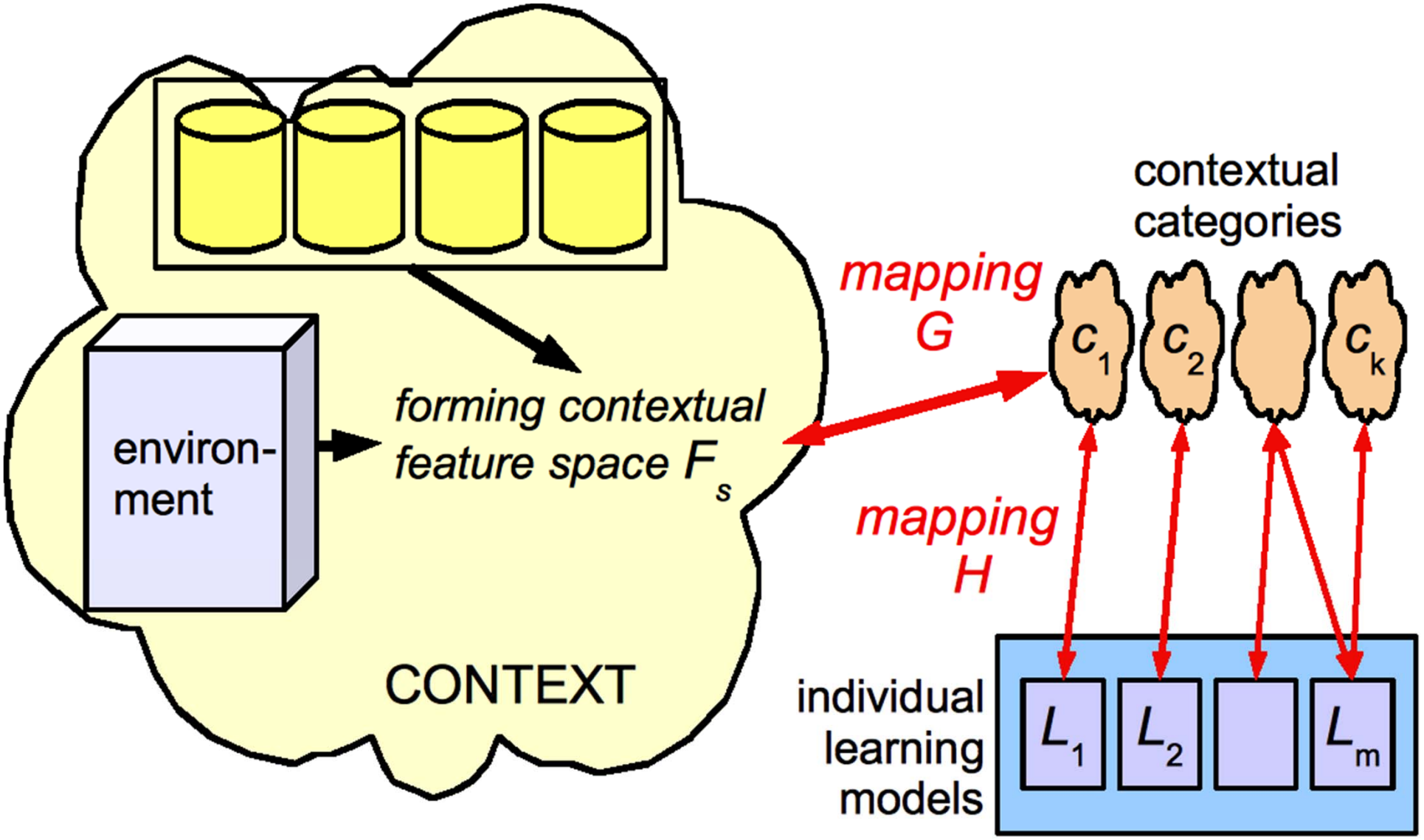
??

Application:

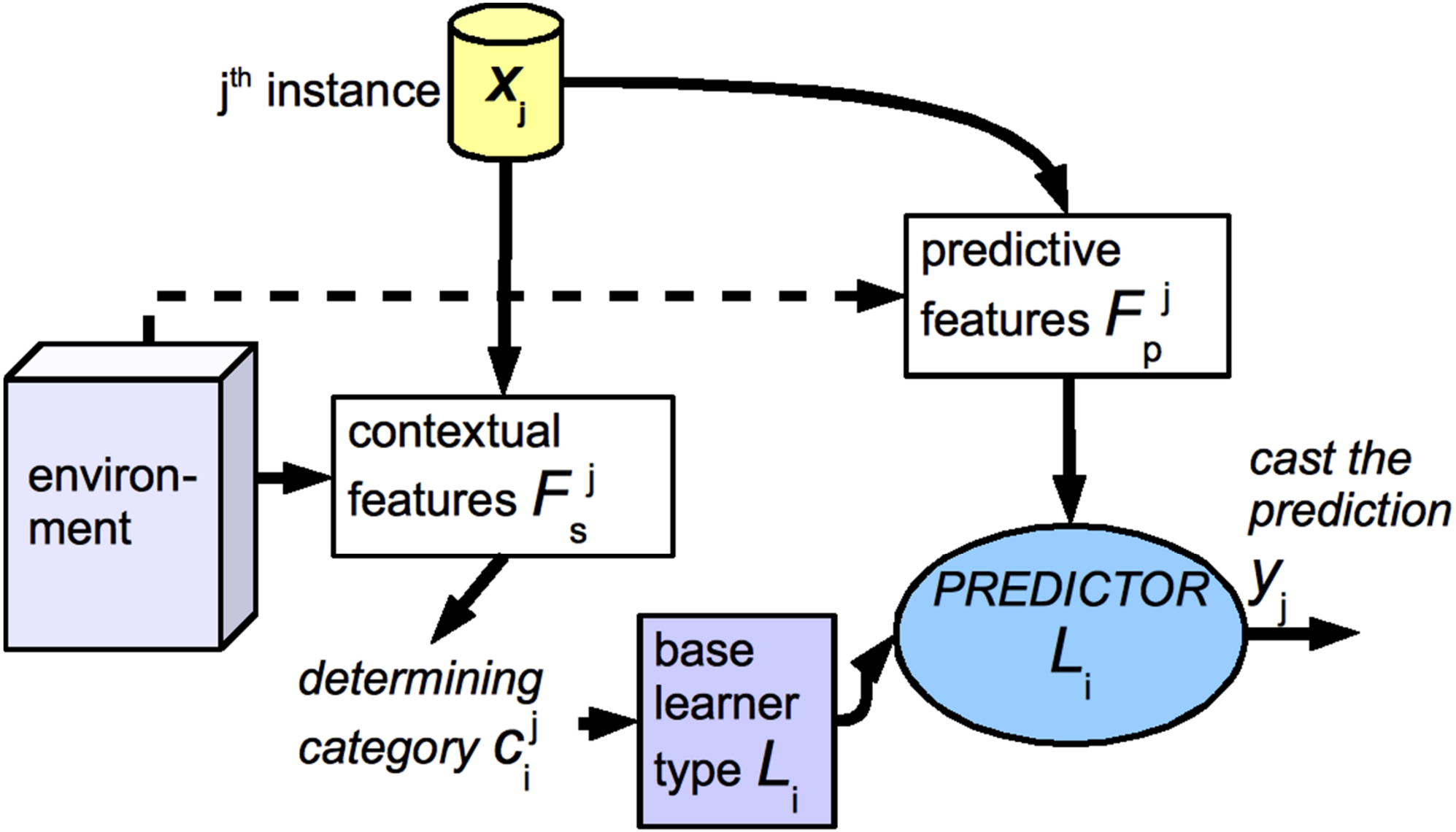
$$y' = L^j(X')$$

$$L^j <= G(X',E)$$

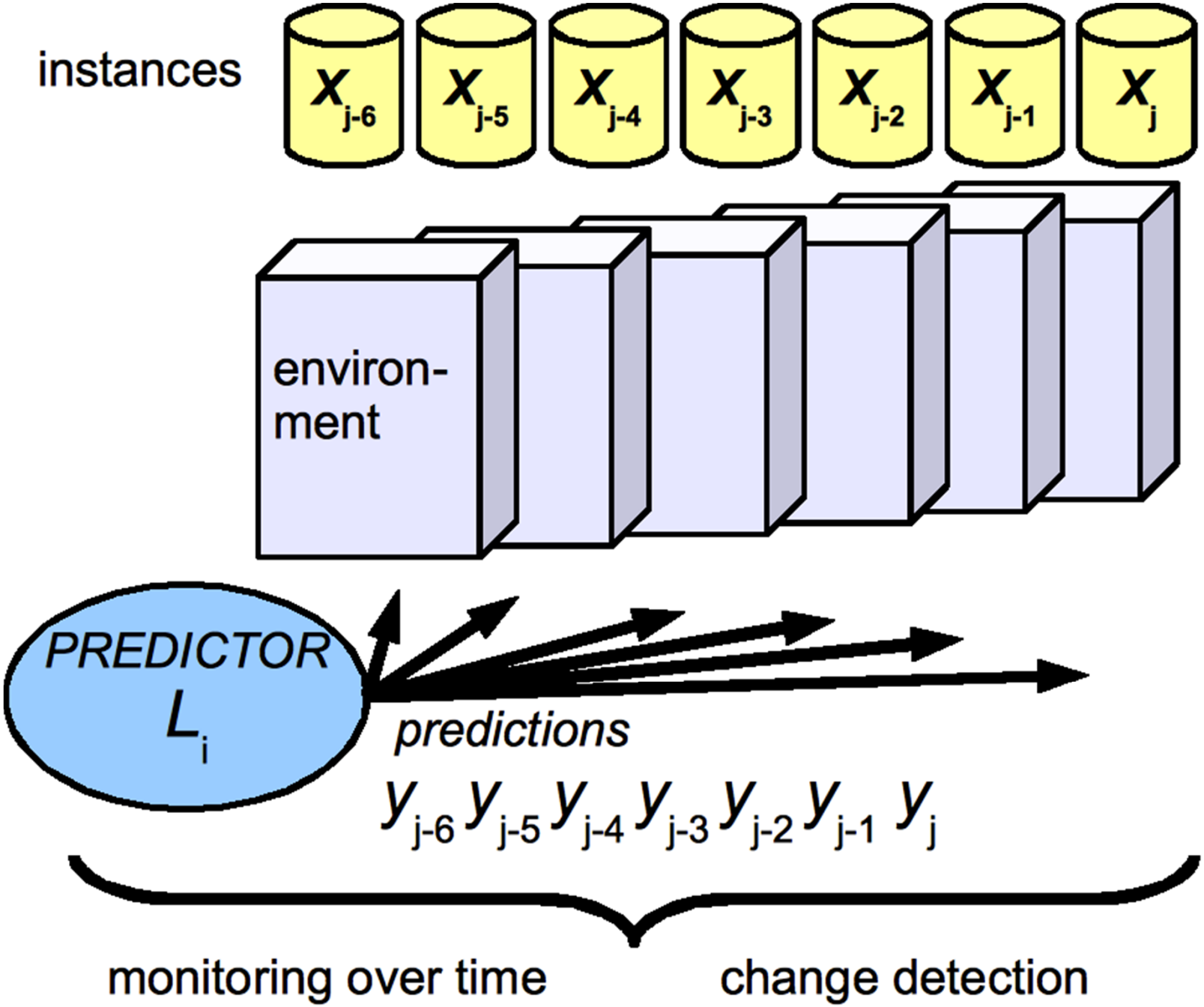
Design: (Re-)Learning Classifiers & Context



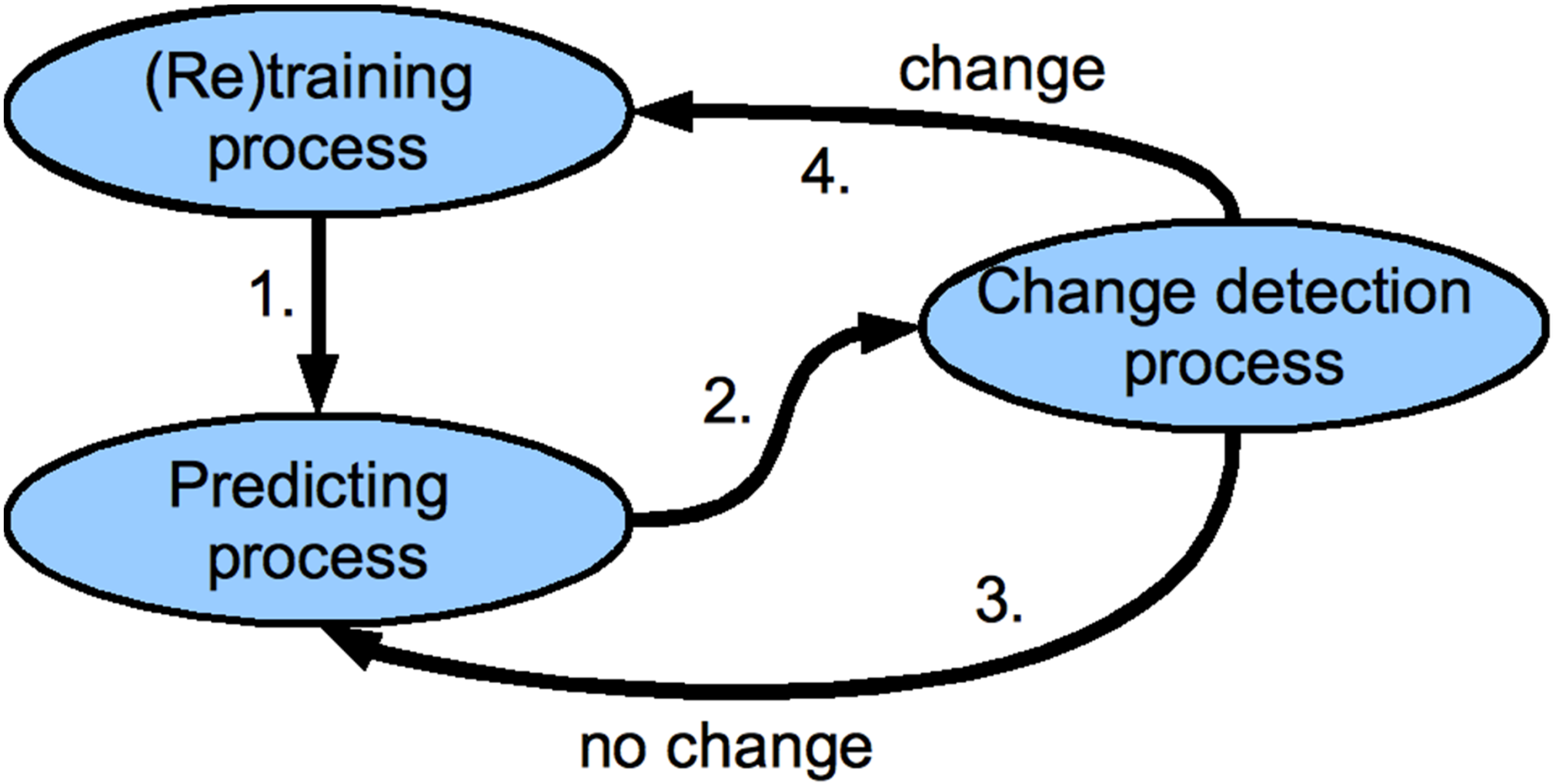
Application: Casting Context-Aware Predictions



Monitoring Changes in Data, Context & Predictions



Three Generic Phases Together



Research Questions

- How to define the context (form and maintain contextual categories) in web analytics?
- How to connect context with the prediction process in predictive web analytics?
- How to integrate change detection mechanisms into the prediction process in web analytics?
- How to ensure integration and feedback mechanisms between change detection and context awareness mechanisms?
- What should a reference architecture allowing to plug in new context aware prediction techniques for a collection of web analytics tasks look like?

STW CAPA: Web Analytics Domain

Our goal: Framework to integrate context–awareness and change detection in Predictive Web Analytics

Two major types of tasks we are interested in

- Demand prediction
- Personalized recommendations

Modeling user behavior based on click–stream data

- Traffic data (page views, visits)
- Navigation data (traversed paths)
- Origin data (referrers, SE, keywords)
- System data (browser, resolution, plugins)

But besides click–stream data

- Customer data (location, lang., history, interests)
- Content data (webpages)

Contextual data (customer related, content related or environment related; calendar, weather, media highlights, twitter stream)

Context in Predictive Web Analytics

- Banner popularity and relevance predictions
  - Clickstream data and banner click data
- Social media and Web Analytics
  - How to link these two sources
- Information portals
  - Bounce rate reduction with recommenders

Framework for Studying Context-aware Banner Selection:

Topical Online Portal Advertising

Agus Wiro Susilo

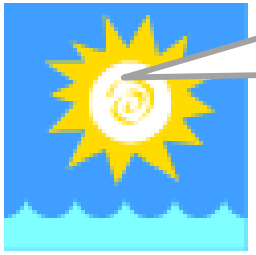
Master project
Feb – Aug 2011
TU Eindhoven, Adversitement, Klikniews

Online advertising settings (stakeholders)

- = user
- = publisher
- = banner
- = ad server
- = advertiser

Context-aware advertising example

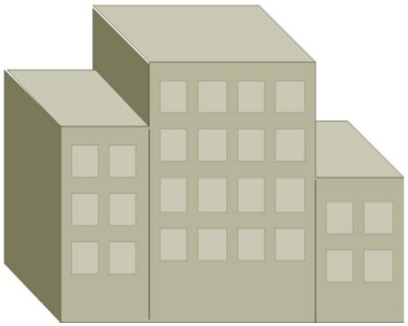
Context:
summer holiday



Two sources for learning context from two separate systems:
- Banner click data
- Clickstream data

More relevant with the context

Context from banner click data



Ad server

- Click time
- User id
- Banner clicked
- etc.

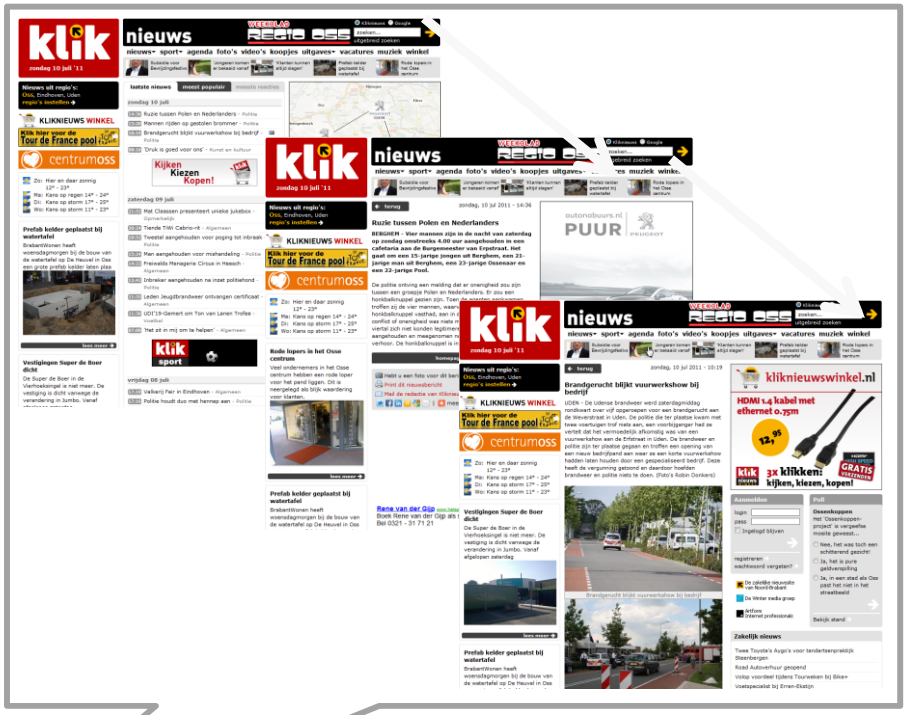
What we **can** learn:

- What and when banner is clicked
- Who clicks the banner

What we **cannot** learn:

- What other banners are shown with the clicked banner
- Clicking user behavior pattern

Context from clickstream data



- Access time
- Page URL
- Banners shown
- IP Address
- User Id
- Screen resolution
- Browser type
- Visit path
- etc.

What we **can** learn:

- User browsing behavior (browsing path)
- User segmentations (by locations, browsers, etc)
- Effectiveness of a page (FAQ, documentation,etc)

What we **cannot** learn:

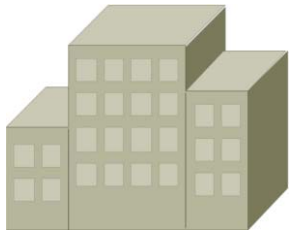
- Banner click pattern (there is no banner click information)

Clickstream + banner click data

Publisher

Clickstream data

Banner click data

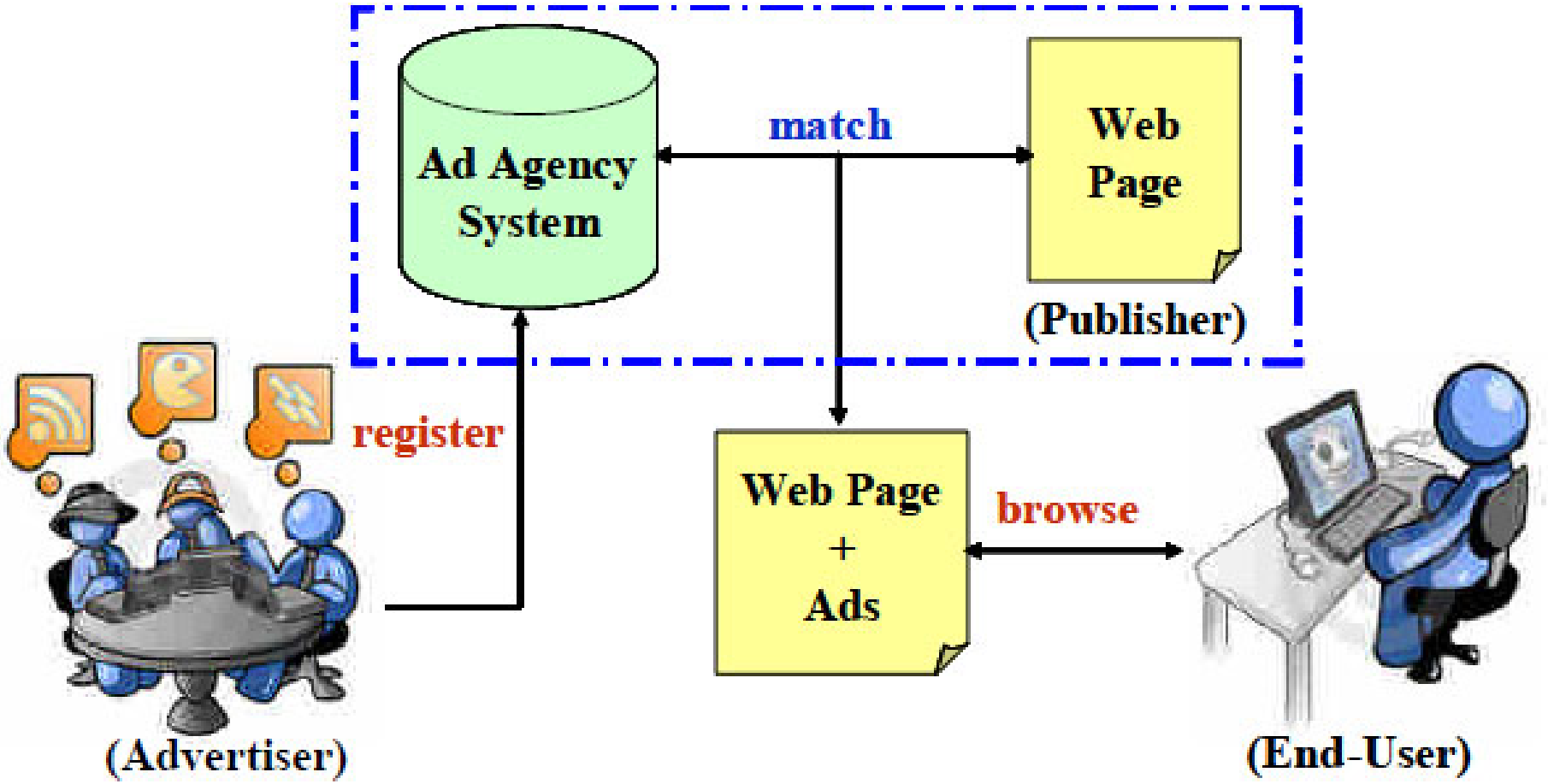


Ad server

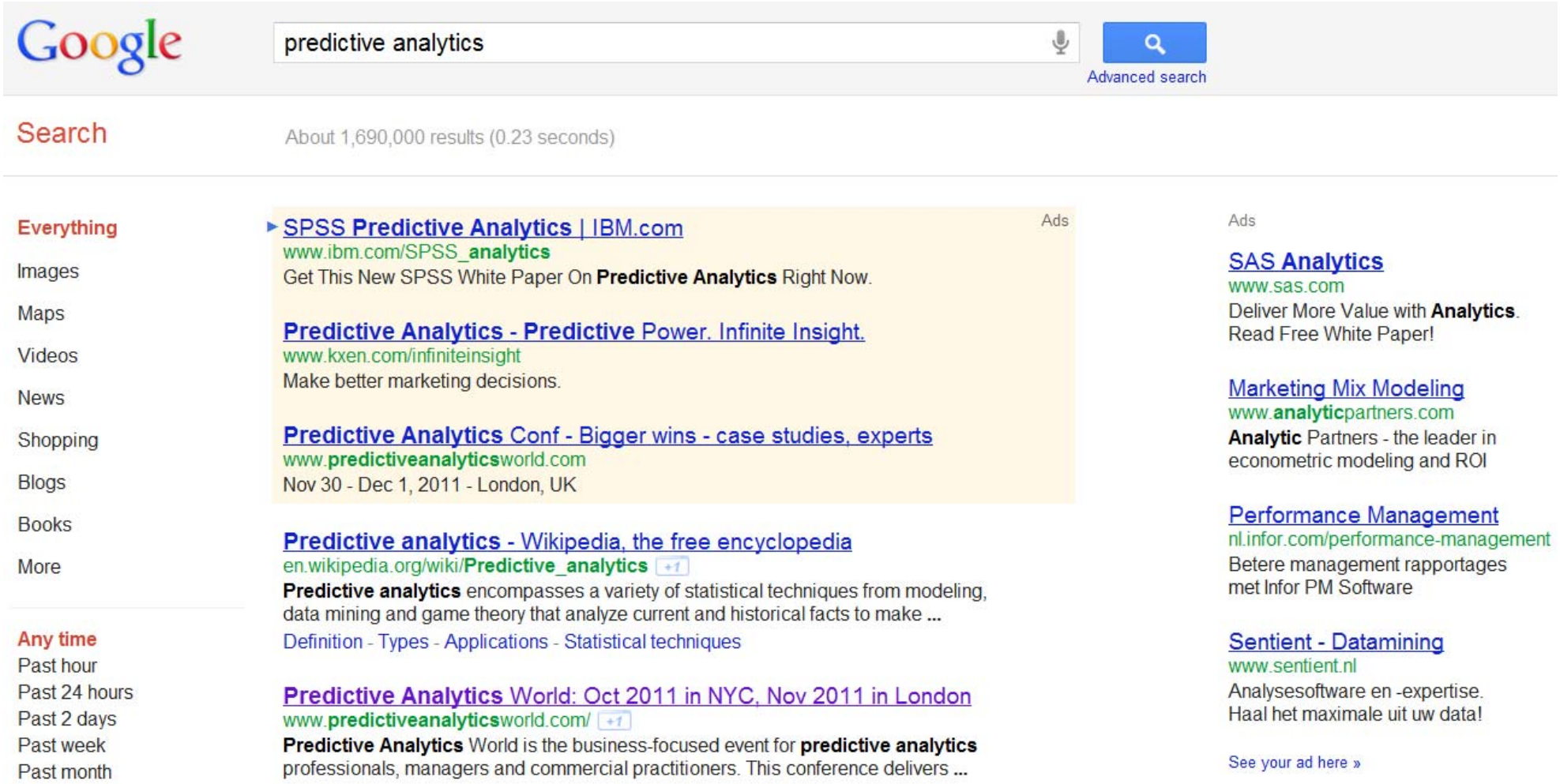
Richer data, from which we **can** learn:

- What banners combinations produce more clicks
- What banners combinations do not produce clicks
- Preferred banners for new user
- Preferred banners for repeating user
- etc.

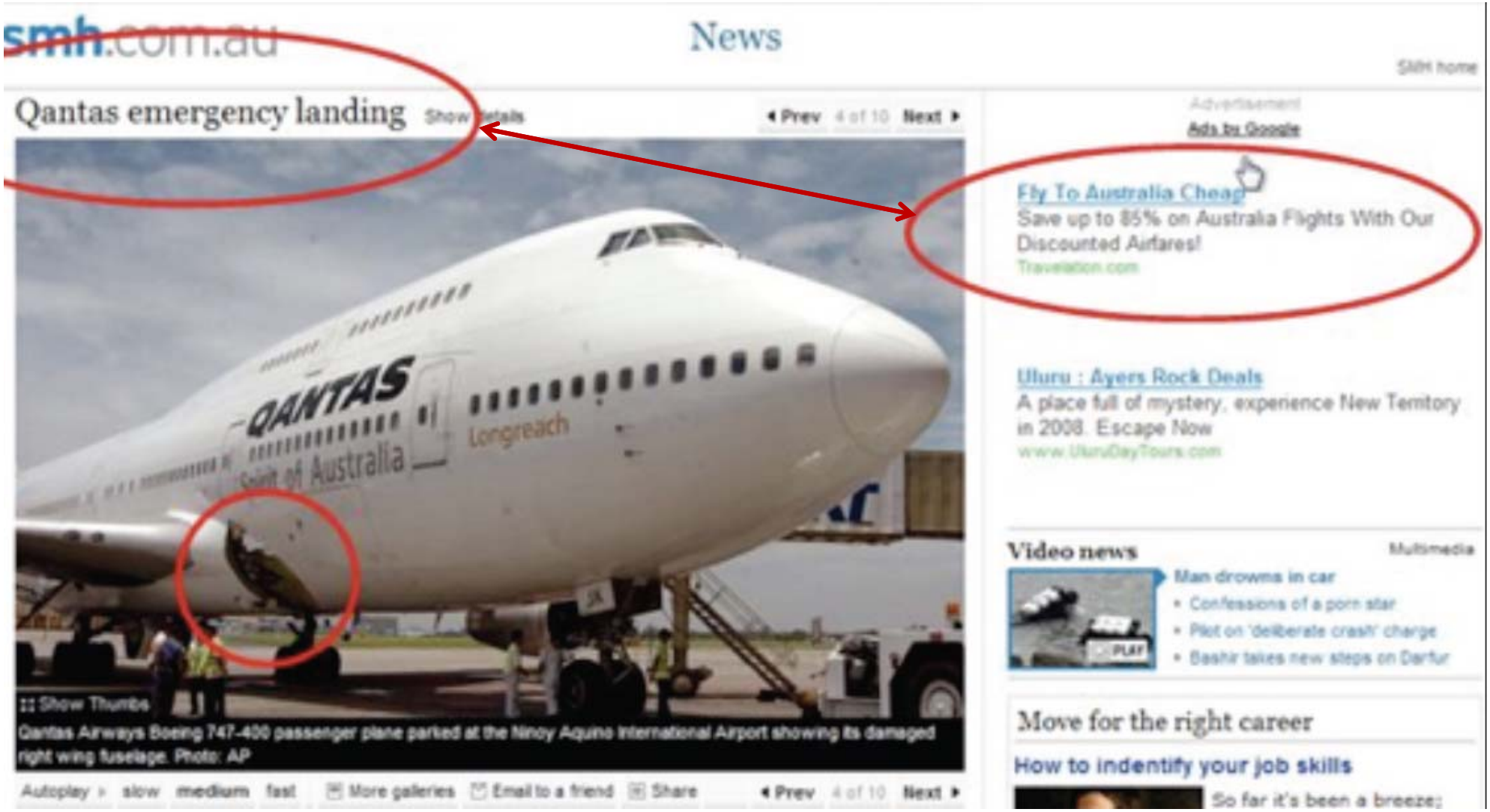
Matching Ads to Content Context



Matching Ads to Content Context



Matching to “Wrong” Context



Matching to “Wrong” Context

Matching to “Wrong” Context

Fox News | Fox Business | Small Business Center | Fox News Radio | Fox News Latino | Fox Nation | Fox News Insider | Register | Login

FOXNEWS Video | Politics | U.S. | Opinion | Entertainment | SciTech | Health | Travel | Leisure | World | Sports | Search

Wildfires Rage in Texas; Thousands Evacuated

Browse Slideshows

Firefighters try to control at least 63 wildfires across the Lone Star state -- including one that has destroyed nearly 600 homes in Central Texas -- as thousands of evacuees take shelter.

Recommend

Image 12 of 18

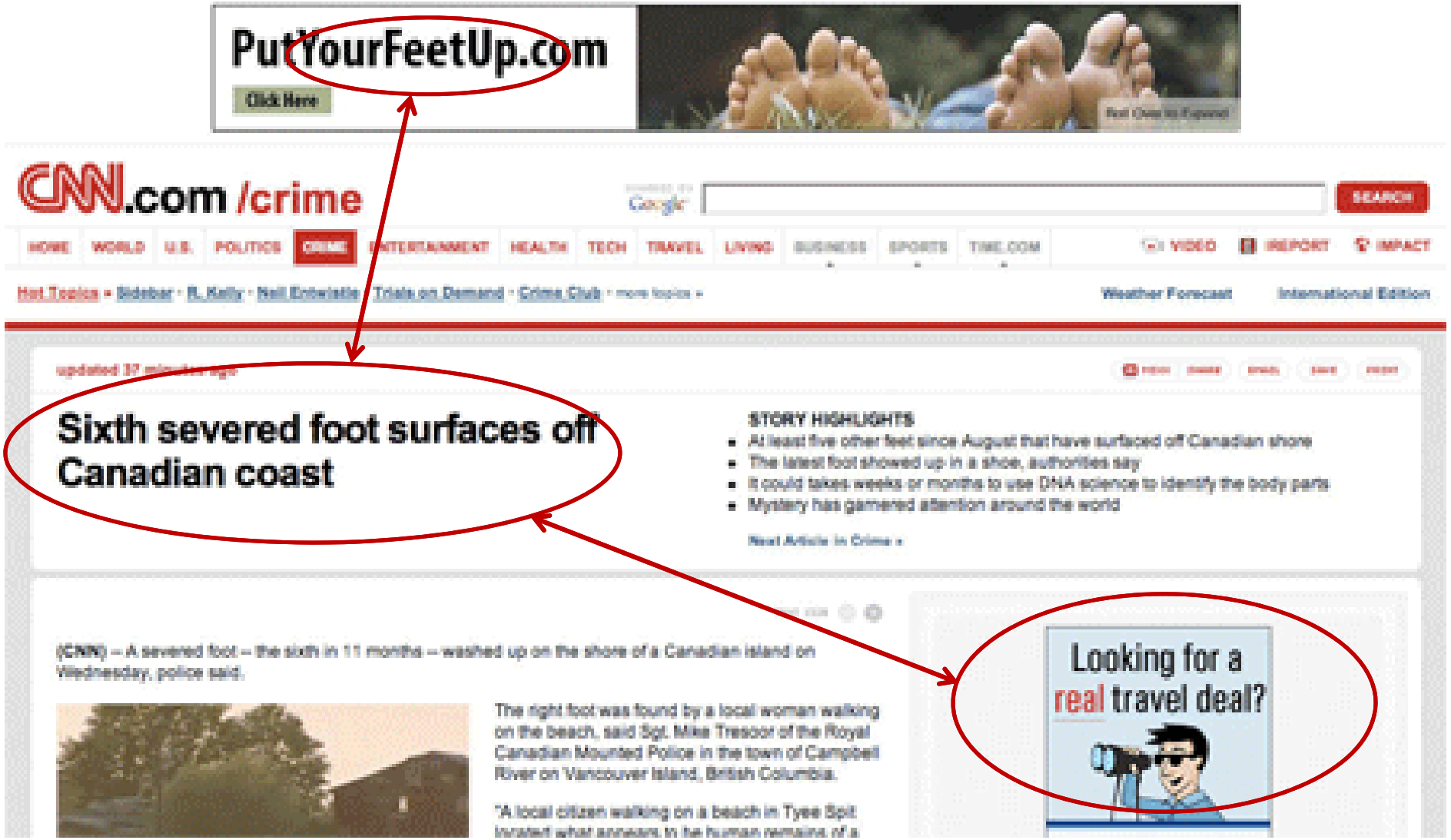
Sept. 6: A Cushing, Texas, firefighter approaches a section of the front line of a massive wildfire near Lilbert, Texas.

Source: AP/The Daily Sentinal

Sponsored Links Buy a link here

Royal Caribbean® Cruises
Reserve A Royal Caribbean® Cruise Today And Find Hot Deals Online!
www.RoyalCaribbean.com

Matching to “Wrong” Context



Social Media Measurement Framework:

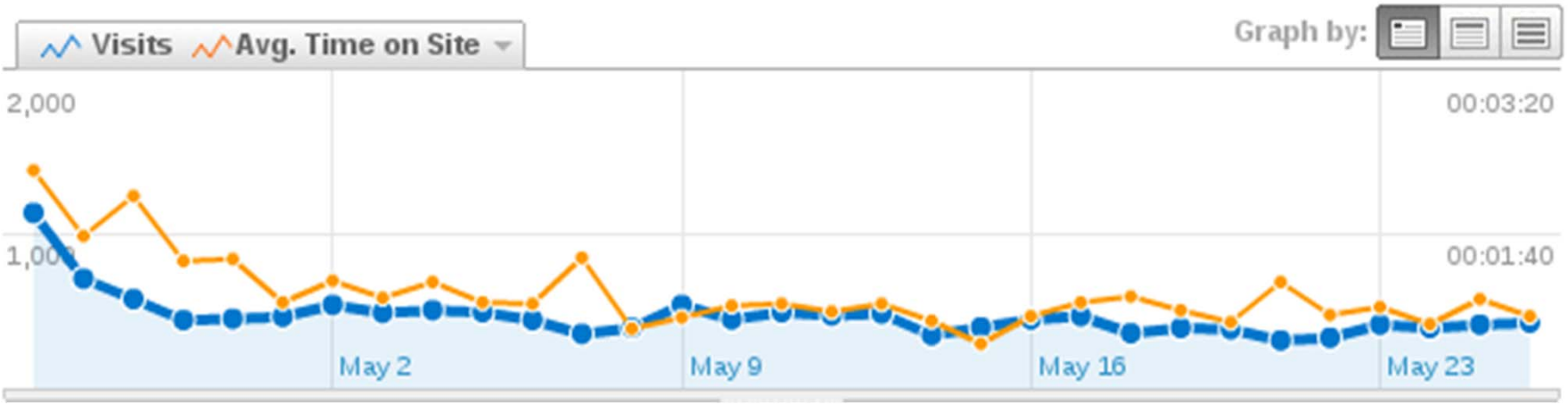
Support for Web Analytics and Search Engine Marketing

Murat Ongan

Master project
Feb – Aug 2011
TU Eindhoven, Adversitement, evaluation: Vodafone data

Web Analytics

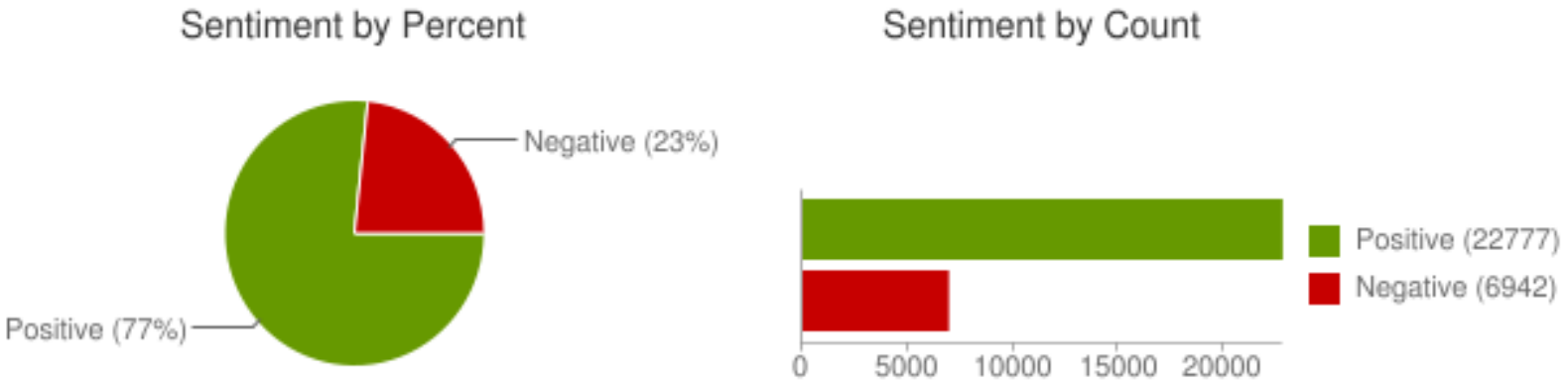
- Measure and report web usage
  - Metrics and dimensions

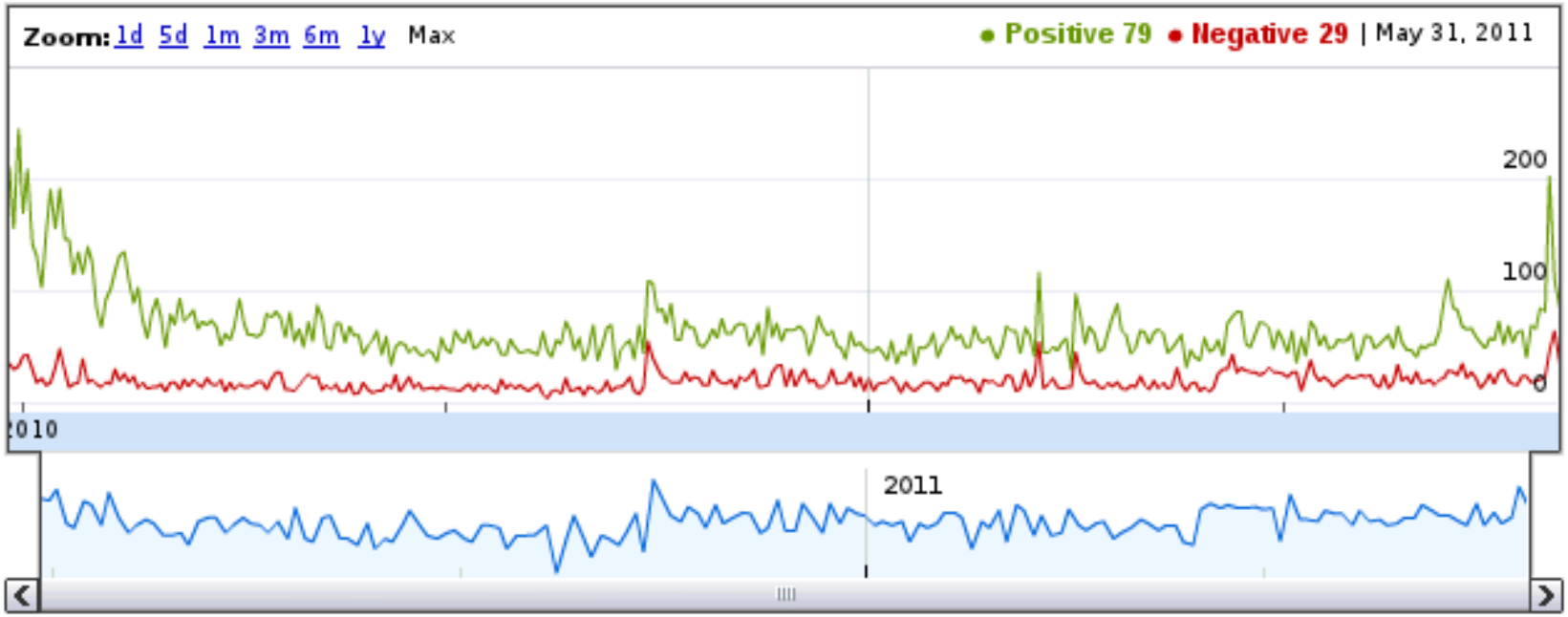


	Source/Medium None	Visits ↓	Pages/Visit	Avg. Time on Site
1.	google / organic	**56,012**	1.59	00:01:03
2.	google.com.tr / referral	**18,577**	1.48	00:00:40
3.	(direct) / (none)	**18,566**	2.37	00:02:03
4.	cclub.metu.edu.tr / referral	**2,305**	3.86	00:03:23
5.	ceng.metu.edu.tr / referral	**1,881**	2.98	00:02:12
6.	facebook.com / referral	**993**	3.16	00:02:33

Social Media Measurement

Sentiment analysis for coca-cola

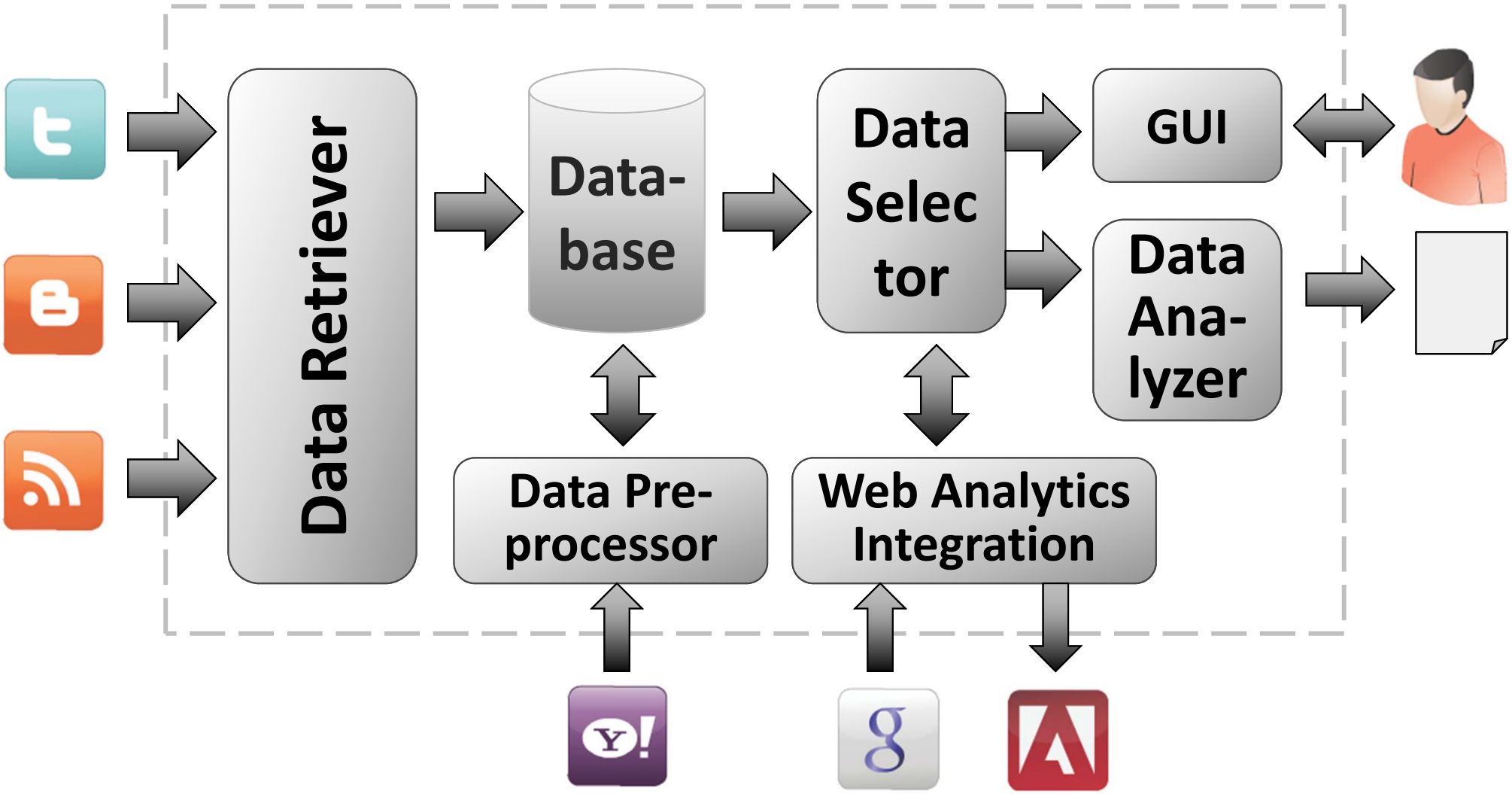




Natural Data Alignment

		Web Analytics	Social Media
Metrics	**Atomic**	Page view	SM post
	Unique	Visitor	Author
	Success	Conversion	Positive message
	Rate	Conversion rate	Percent positive
Dimensions	**Date/time**	Yes	Yes
	Location	Yes	Yes
	Medium	Direct, organic, referral	Twitter, news, blog
	Source	Website domain	Twitter user / domain
	Feature	Search keyword	Post content

Architecture Overview



mastersportal.eu

TU/e Technische Universiteit Eindhoven University of Technology

(Context-aware)

Retrieval & Recommendation of Higher Education Programmes:

I Know Where You Will Study Next Year

Thijs Putman

Master project
Oct 2009 – May 2010
TU Eindhoven, MastersPortal

Thesis is accessible at http://www.win.tue.nl/~mpechen/projects/pdfs/Putman2010.pdf

Search

Discipline

and / or

business information systems

Advanced Search

Refine search

Your search criteria

Netherlands

Language of Tuition

Set

Region

Greater Europe

Any country Set

Start Semester

Spring
Autumn

Educational Form

Taught
Research

Alternatives

Part-time
Full-time
Evening
Self-study

Found 5 programmes, displaying 1 - 5

› Netherlands (5)

Name | Tuition Fee | Duration | 10 results | No grouping

Business Information Technology (MSc)

University of Twente (UT) – Enschede, Netherlands

Duration: 24 months
Annual Tuition Fee: € 1,672

UNIVERSITY OF TWENTE.

As a Business Information Technology professional, you know that introducing a new information system is more than a technical installation, and requires insight in business processes and culture.

Detailed Description

Business Information Systems (MSc)

Eindhoven University of Technology (TU/e) – Eindhoven, Netherlands

Duration: 24 months
Annual Tuition Fee: € 1,620

The Master's degree program in Business Information Systems combines computer science and business management. The program places the emphasis on the development of high-quality information systems based on a business perspective.

Detailed Description

IMMIT: International Master in Management of Information Technology,... (MSc)

Tilburg University – Tilburg, Netherlands

Duration: 20 months
Annual Tuition Fee: € 4,000

Information and information technology (IT) are increasingly an integral part of products and services as well as the foundation of business processes. Organisations must know how to make the right choices with respect to new IT systems and at the...

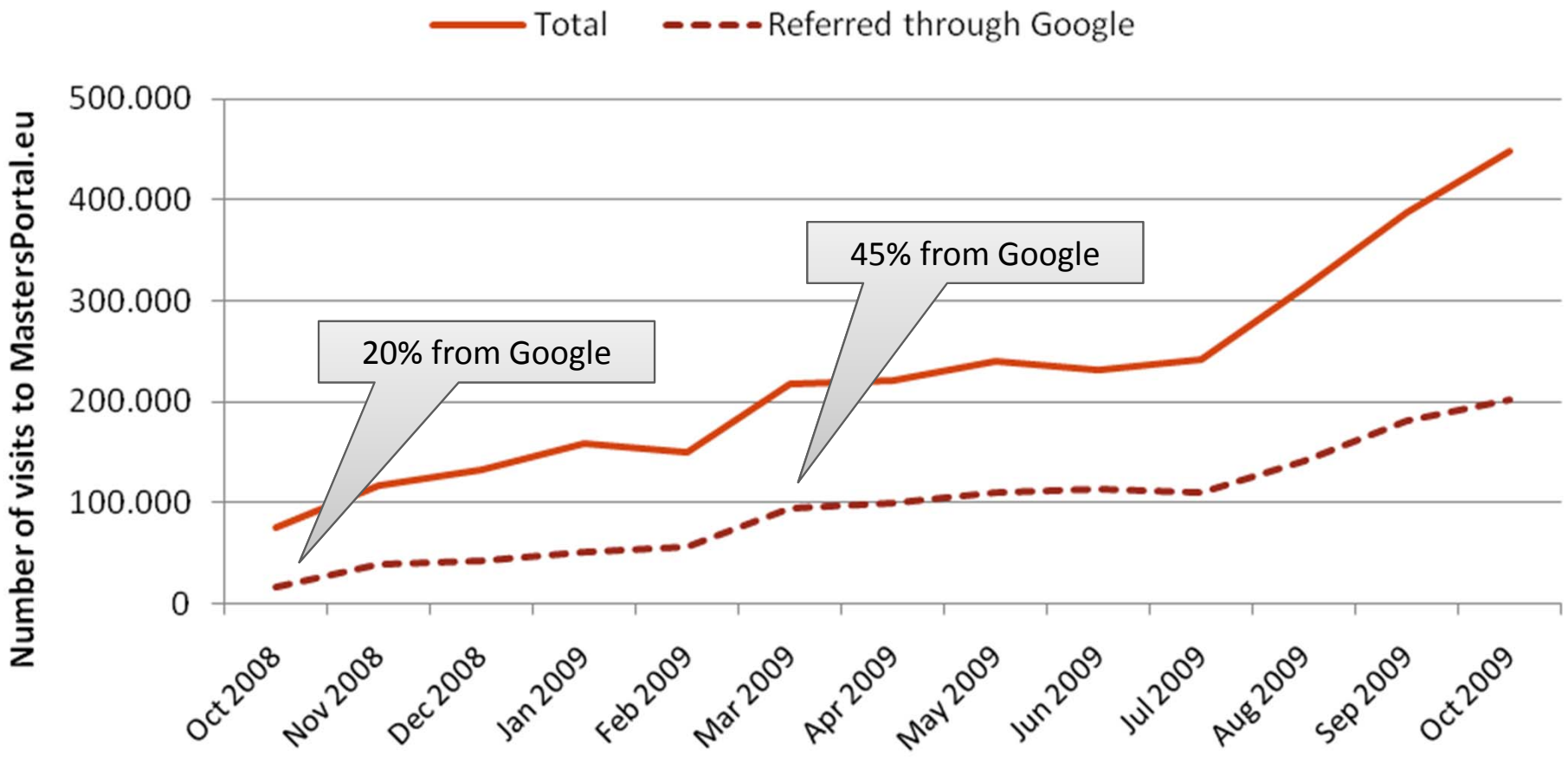
Detailed Description

IMSE: International Master in Service Engineering, Erasmus Mundus (MSc)

Tilburg University – Tilburg, Netherlands

Duration: 24 months
Annual Tuition Fee: € 4,000

Impact of Google on MastersPortal.eu

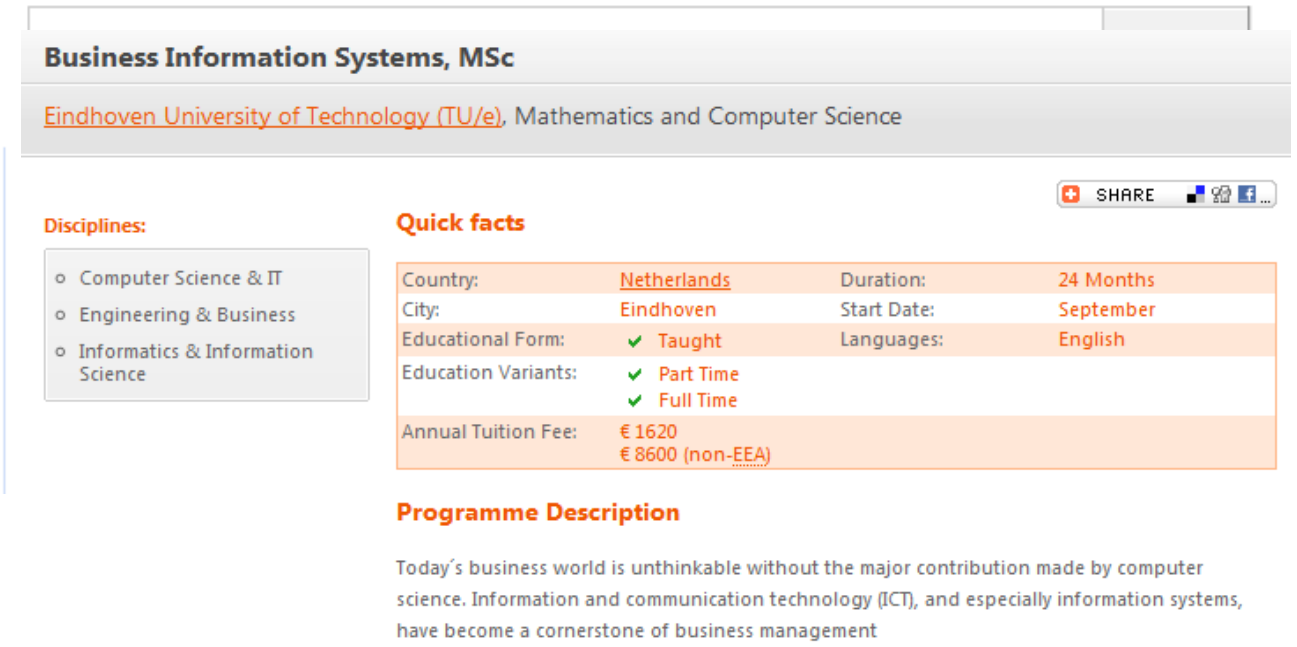


- Google's share steady **around 50%** since March 2009

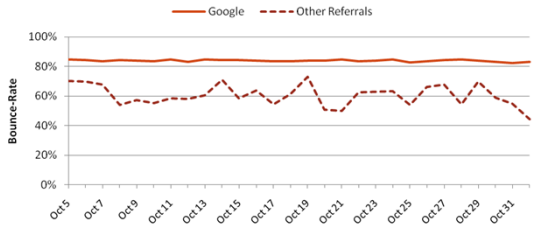
Impact of Google on MastersPortal.eu

- Visitors are often **referred** to pages with **detailed info**

Google



- Many visitors leave after **first page**
  - Leads to a high **bounce-rate**



    - Bounce-rate for Google **84%;** Other, non search-engine, referrals: **65%**

Solution: Programme Recommender

- Provide a set of
  "Related Master's Programmes"
- Based on the **academic contents** of each Master's programme
  - Select from the over **14.000 programmes** in the database
  - Mostly **unstructured** information

Search

Discipline

and / or

business information systems

Advanced Search

Related Programmes

- Master in Applied Economic Sciences: Business Engineering in Management Information Systems (MSc) — Hasselt University
- Information and Service Management (MSc) — Helsinki School of Economics
- Information Management (MSc) — Catholic University of Leuven
- Information Systems (MSc) — University of Sheffield
- Computer Science and Engineering (MSc) — Eindhoven University of Technology
- Information Systems and Technology (MSc) — City University London
- Information Technology (MSc) — Bournemouth University
- Information Technology and Management (MSc) — Jönköping University

What is this?

Business Information Systems, MSc

Eindhoven University of Technology (TU/e), Mathematics and Computer Science

Disciplines:

- Computer Science & IT
- Engineering & Business
- Informatics & Information Science

Quick facts

Country:	Netherlands
City:	Eindhoven
Duration:	24 Months
Starting Date:	September
Education Variants:	✔ Part Time ✔ Full Time
Educational Form:	✔ Taught
Languages	English
Annual Tuition Fee:	€ 1620 € 8600 (non-EEA)

Programme Description

Today´s business world is unthinkable without the major contribution made by computer science. Information and communication technology (ICT), and especially information systems, have become a cornerstone of business management
in multinationals, in banks and insurance companies, and in small and medium-size enterprises. Companies have become dependent on these increasingly complex systems, and out of necessity place stringent demands on their reliability and security.

The Master´s degree program in Business Information Systems combines computer science and business management. The program places the emphasis on the development of high-quality information systems based on a business perspective.

As a graduate of this program you will combine a scientific attitude with a model-driven engineering approach. You will be able to understand the demands that are placed on information systems, and to initiate and implement new applications. This approach can already be seen in the compulsory courses of the program. The compulsory computer science courses are Software architecting, Web information systems, Database models, Process modeling and Information retrieval. The compulsory business management courses are Information management, IT governance, E-business architecture and systems, Workflow management systems and Supply chain logistics and information management.

During the program you can place the emphasis on the computer science aspects or the business

TAMPERE UNIVERSITY OF TECHNOLOGY
Study Engineering in Tampere Finland

IUM INTERNATIONAL UNIVERSITY OF MONACO
Masters
FINANCE
LUXURY GOODS & SERVICES
INTERNATIONAL BUSINESS & GLOBAL AFFAIRS

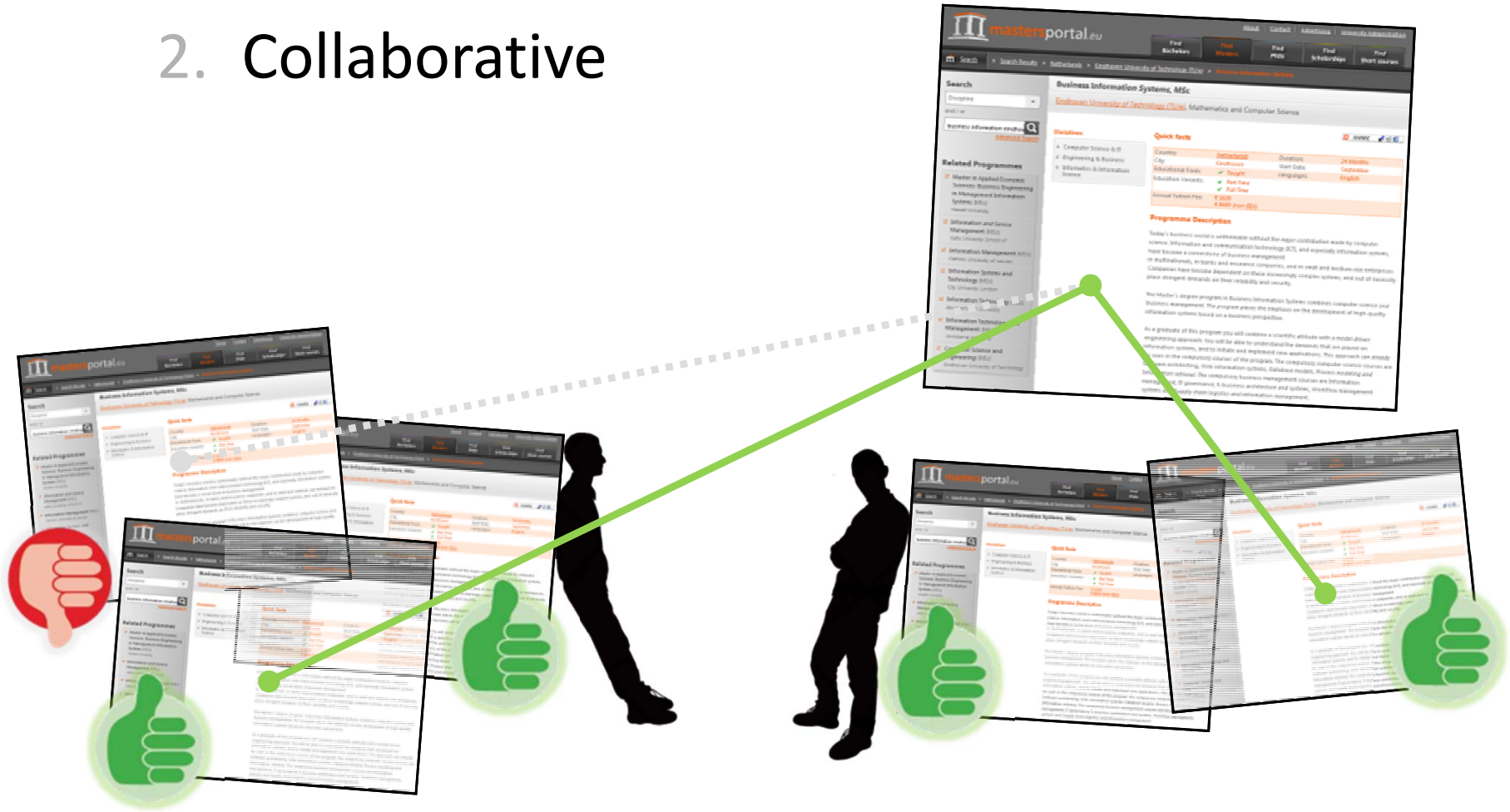
One Year Full-Time
Int. Master (MSc/MA)
TiasNimbas Business School

Deploy & Evaluate Two Alternatives

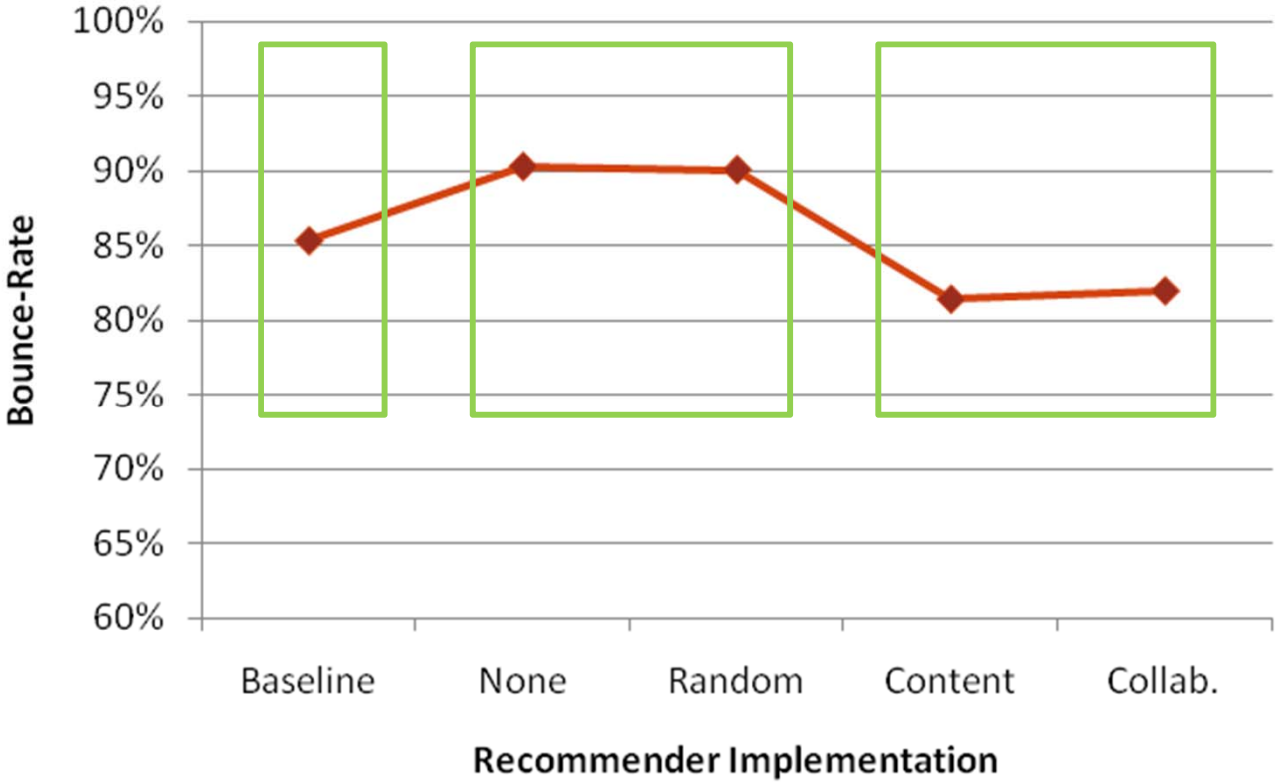
- *state-of-the-art* **recommendation approaches**
  1. Content-Based

Deploy & evaluate two alternatives

- *state-of-the-art* **recommendation approaches**
  1. Content-Based
  2. Collaborative



Experiment Results of Live Testing



	Bounce-rate *
None	90,29%
Random	90,08%
Baseline	85,43%
Collaborative	82,03%
Content	81,49%

- Both approaches perform equally well
- Almost **doubles** visitor retention

Contextual Factors

- Categorising the visitor
  - Create groups of similar visitors; **visitor segments**
- Is performance **consistent** across different **visitor segments**?

An Adaptive Recommender System

Content-Based Recommender

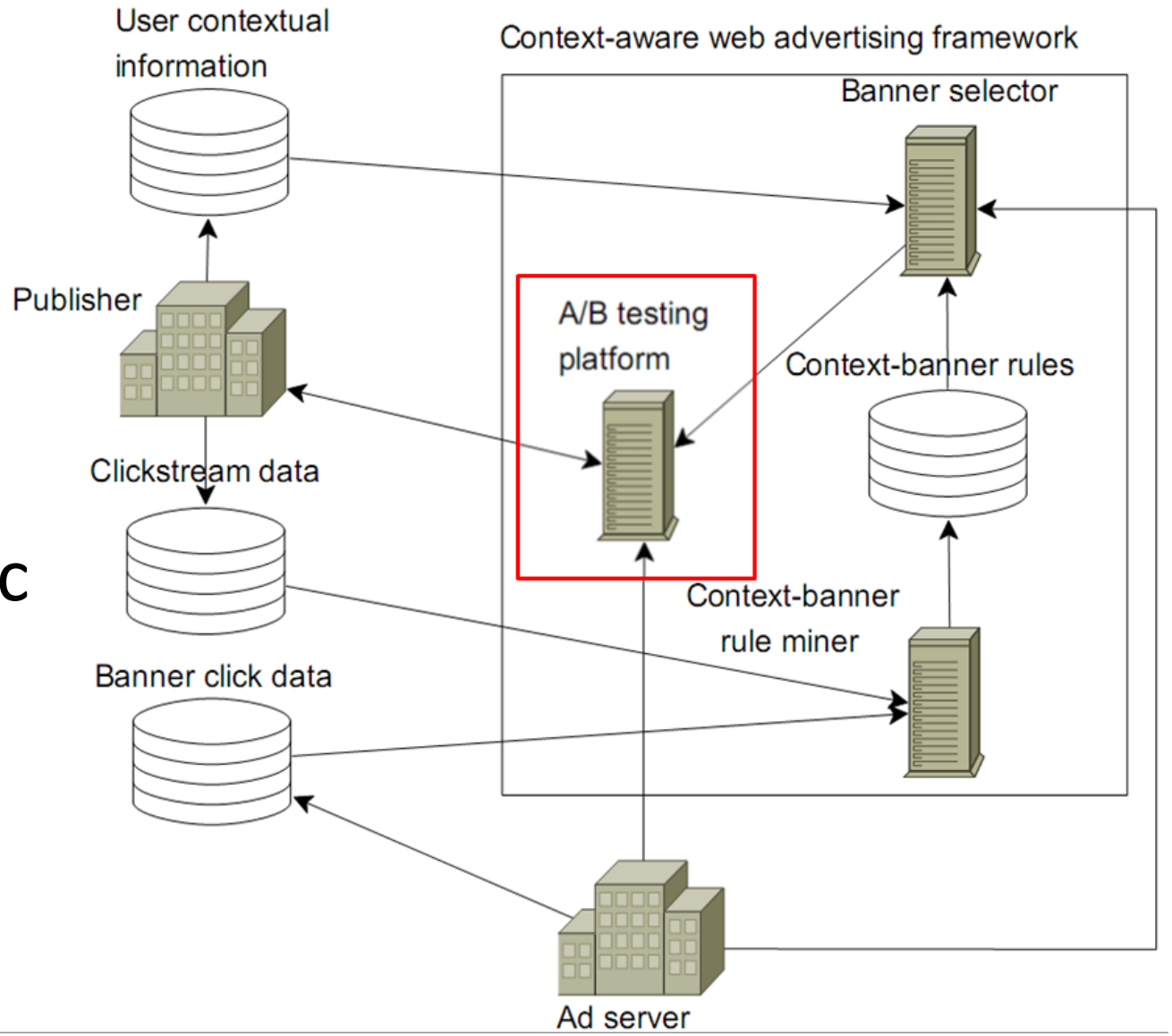
Collaborative Recommender

A/B Live Testing

- An important element of R&D infrastructure
  - MastersPortal
  - Klicknews
  - Adversitement

  Web Analytics solution
- Two versions of the system

- Splitting the traffic between them
- Compute which one is doing better



Planned and Ongoing Work

Together with Adversitement & Klikniews

- Finding (anti)catalyst banners
  - Sports fits politics, but not fashion
- Study A/B testing infrastructure
  - experiment design, mining the results from the controlled experiments
- Matching Ads to Content Context
  - Also wrt positive–negative content
- Analysing different problems formulations: e.g. contract items as constraints (number of impressions to be guaranteed)

CAPA Research Team

? PhD

? Postdoc