DECIBELdigital

LONDON
metropolitan
university

Enhancing online analytics tools with dynamic user profiling using intelligent methods for classification

Dr. Vassil T. Vassilev

School of Computing

Disclaimer

LONDON
metropolitan
university

- The experimental research and development work reported here has been carried out during the period 2010-2013 in a collaborative project KTP008263 between **London Metropolitan University** and **Decibel Digital** (former New Brand Vision) under the join supervision of **Prof. Dominic Palmer-Brown**, *Dean of the Faculty of Life Sciences and Computing* of Londonmet and **Ben Harris**, *President of Decibel Digital* group of companies.
- The project has been jointly funded by the **Technology Strategy Board** of UK (TSB) and Decibel Digital in the framework of the Knowledge Transfer Programme of TSB (KTP). On behalf of Londonmet the project has been managed by **Dr. Vassil T. Vassilev** and on behalf of Decibel Digital - by **Timothy De Paris**.
- The results of the work on the project are embedded into the commercial software products of the company **Decibel Insight©** and are entirely owned by the company.

Content

LONDON
metropolitan
university

1 **The Problem**

2 **The Standard Solution**

3 **Methods for Online Classification and Clickstream Data Analysis**

- ◆ Heatmaps for Visualisation and Analysis of User Interaction and Navigation
- ◆ Neural Networks for Fast Classification of Customer Data
- ◆ Linguistic Thesaurus for Disambiguation of Search Terms

4 **Further Development**

1 The Problem

- **Dynamic User Profiling**
  - How to capture and further analyze the online user behavior without interfering with the business operation
- **Applicability of the Dynamic User Profiling**
  - Individual Identification based on tracking user interactions for security management
  - Pattern recognition based on reconstruction of user journeys for fraud detection
  - Customer classification based on preferences for market segmentation
  - Keyword analysis for Search Engine Optimization
  - Keystroke, mouse and touch screen dynamics analysis for Web Design Optimization

Typical online datasets which contain relevant information

LONDON metropolitan university

Timing of the journey (absolute and relative start and end time, duration of the time on a page, duration of the overal site journey, etc.)

Origin of the journey (network, country, computer, browser, etc.)

User Interaction habits (scrolling, dragging, mouse movements, key holding, window resizing, etc.)

User Exploration habits (page sequencing, returns and repeats, etc.)

User Navigation habits (selecting from drop-down menus, clicking on hyperlinks, pushing buttons, etc.)

User Lifecycle habits (starting, drilling down, transactions completion, etc.)

Why the dynamic construction of the user profile can be very difficult?

- Dependence on external factors
  - Often the methods rely on running scripts on the client (typically JavaScript), which may be disabled
  - The session control is of paramount importance for the operation, but the sessions can also be unavailable
- Large volume of online data
  - The volume of clickstream data can be extremely big (i.e., 60 TB data for 1 month in a typical transactional system with several millions of online users) which completely rules out offline and batch mode online data analysis
  - The time for filtering and pre-processing becomes comparable with the collection time, which completely rules out many methods for online analysis

2 The Standard Solution

- Install a plugin on the corporate Web server which captures the user interactions with the Web site and/or monitors the traffic to capture the user navigation on the Web pages
- Upload the captured data to a third-party server equipped with data analysis tools to perform external data analysis
- Performing (offline) data analysis and periodic report generation

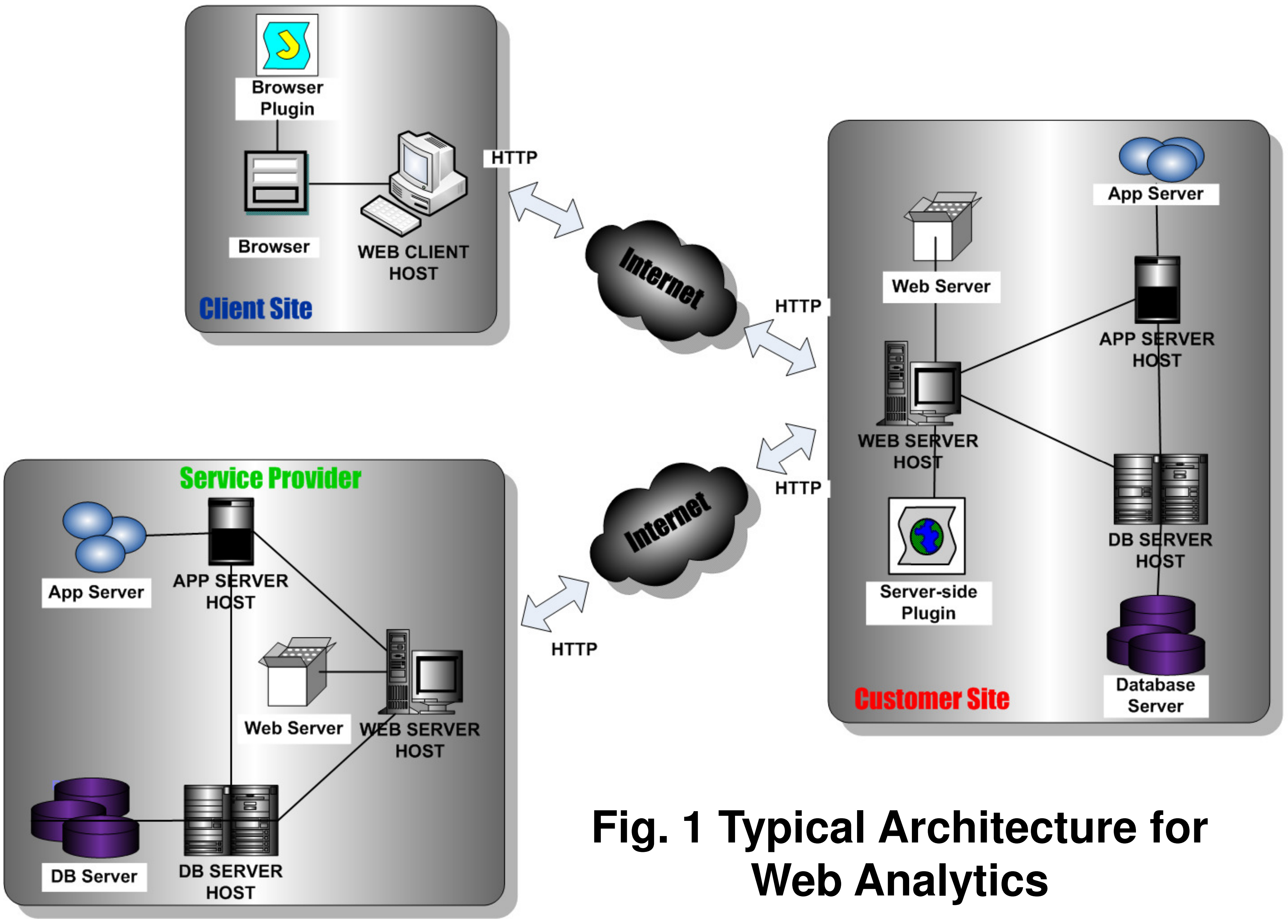


Fig. 1 Typical Architecture for Web Analytics

Recent Champions of Online Analytics

BioCatch: Fraud Detection using Biometric Methods with application to e-Commerce and online banking. Supports both Web and Mobile Web analytics. Captures the data on the client in real time and provides the analytics services on the cloud.

Webtrends Analytics: Market Segmentation for e-Commerce. Supports Web analytics only. Analyzes the data from the server logs in batch mode and provides the analytics services on a corporate server.

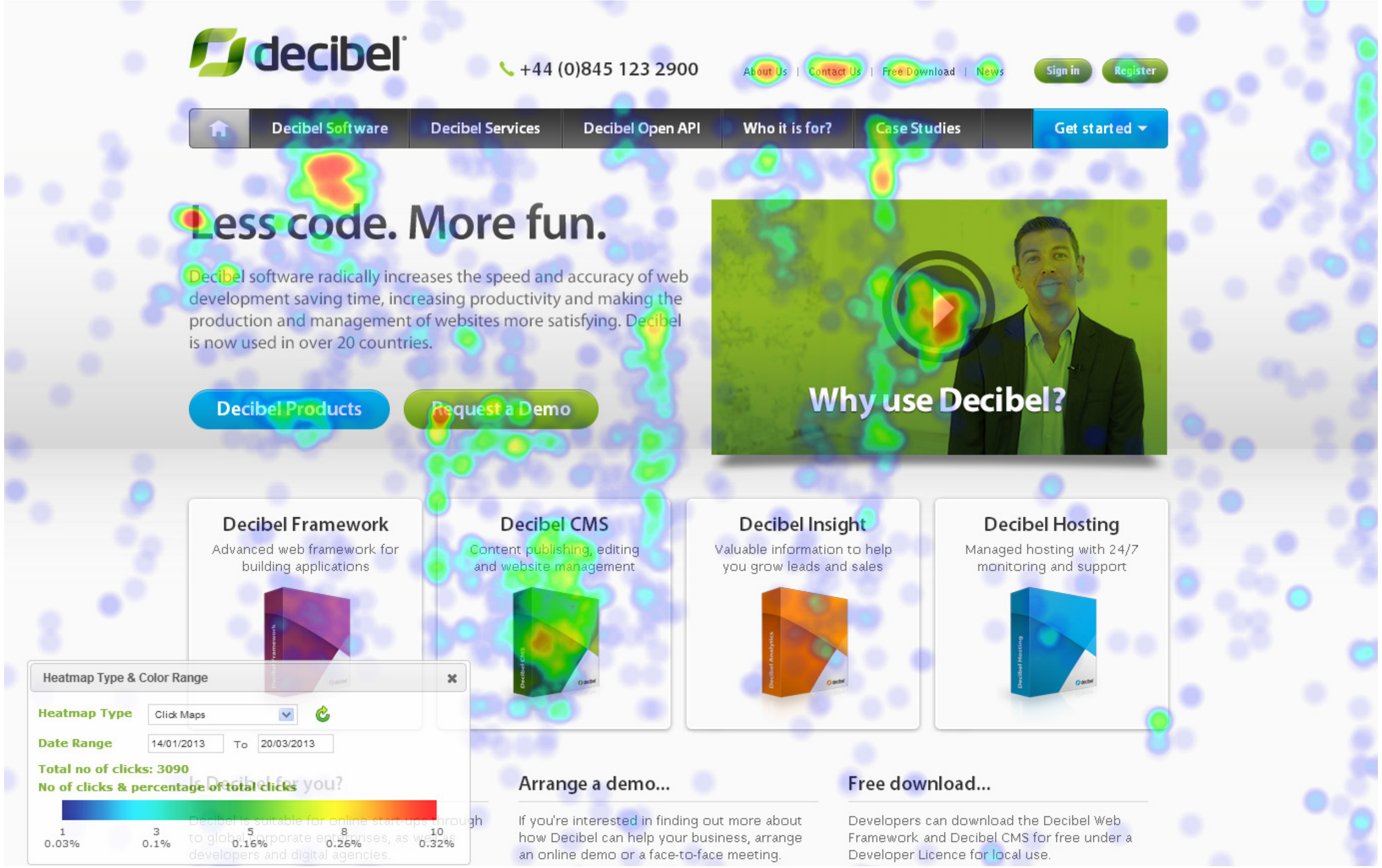
Tealeaf: User tracking for Customer Management in e-Commerce and office. Captures the entire user behavior on the client, which is then played on the server for further analysis and provides the analytics services on a corporate server or on the cloud.

3 Methods for Online Classification and Clickstream Data Analysis

3.1 Heatmaps

- Used for visualization of the user interaction with the Web page for the purpose of Web Design Optimization [1]
- Based on the use of JavaScript API, which is specific for each browser
- Typically visualized in a 2D plane (*pixel map*) but can be mapped to a component container (*symbolic map*) and can be also combined with additional attributes *(value map)*
- Fully supported by **Decibel Insight** v. 2

Heatmap of the user clicks on a Web Page



Less code. More fun.

Decibel software radically increases the speed and accuracy of web development saving time, increasing productivity and making the production and management of websites more satisfying. Decibel is now used in over 20 countries.

Decibel Framework

Advanced web framework for building applications

Decibel CMS

Content publishing, editing and website management

Decibel Insight

Valuable information to help you grow leads and sales

Decibel Hosting

Managed hosting with 24/7 monitoring and support

Is Decibel for you?

Decibel is suitable for online start-ups through to global corporate enterprises, as well as developers and digital agencies.

Arrange a demo...

If you're interested in finding out more about how Decibel can help your business, arrange an online demo or a face-to-face meeting.

Free download...

Developers can download the Decibel Web Framework and Decibel CMS for free under a Developer Licence for local use.

Heatmap Type & Color Range

Heatmap Type: Click Maps

Date Range: 14/01/2013 To 20/03/2013

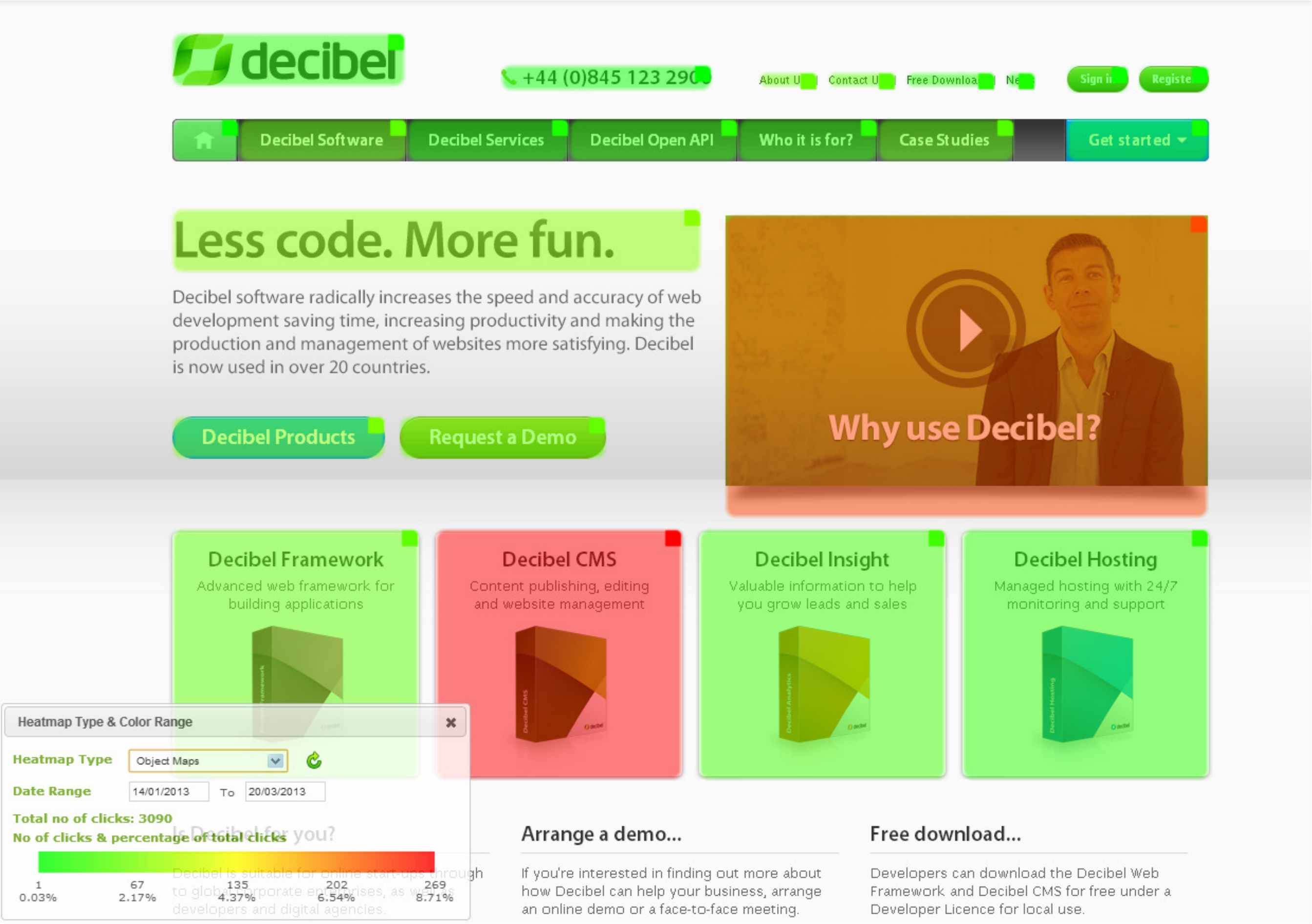
Total no of clicks: 3090

No of clicks & percentage of total clicks

1	3	5	8	10
0.03%	0.1%	0.16%	0.26%	0.32%

Value heatmap of the Web page areas

LONDON
metropolitan
university



Less code. More fun.

Decibel software radically increases the speed and accuracy of web development saving time, increasing productivity and making the production and management of websites more satisfying. Decibel is now used in over 20 countries.

Decibel Framework
Advanced web framework for building applications

Decibel CMS
Content publishing, editing and website management

Decibel Insight
Valuable information to help you grow leads and sales

Decibel Hosting
Managed hosting with 24/7 monitoring and support

Is Decibel for you?
Decibel is suitable for online start-ups through to global corporate enterprises, as well as developers and digital agencies.

Arrange a demo...
If you're interested in finding out more about how Decibel can help your business, arrange an online demo or a face-to-face meeting.

Free download...
Developers can download the Decibel Web Framework and Decibel CMS for free under a Developer Licence for local use.

Heatmap Type & Color Range ✖

Heatmap Type: Object Maps

Date Range: 14/01/2013 To 20/03/2013

Total no of clicks: 3090

No of clicks & percentage of total clicks

1	67	135	202	269
0.03%	2.17%	4.37%	6.54%	8.71%

3.2 Neural Networks for Online Data Classification

- Used for classification of the user behavior based on a set of pre-formatted attributes of the URLs for the purpose of market segmentation
- Based on a special class of neural networks with fast algorithm for unsupervised learning, developed by **Prof. Dominic Palmer-Brown** and his co-workers over the last 10 years (*Snap-Drift Neural Networks*) [2]
- Implemented experimentally as an offline batch processing utility and planned to be incorporated into the next version of **Decibel Insight** v. 3.

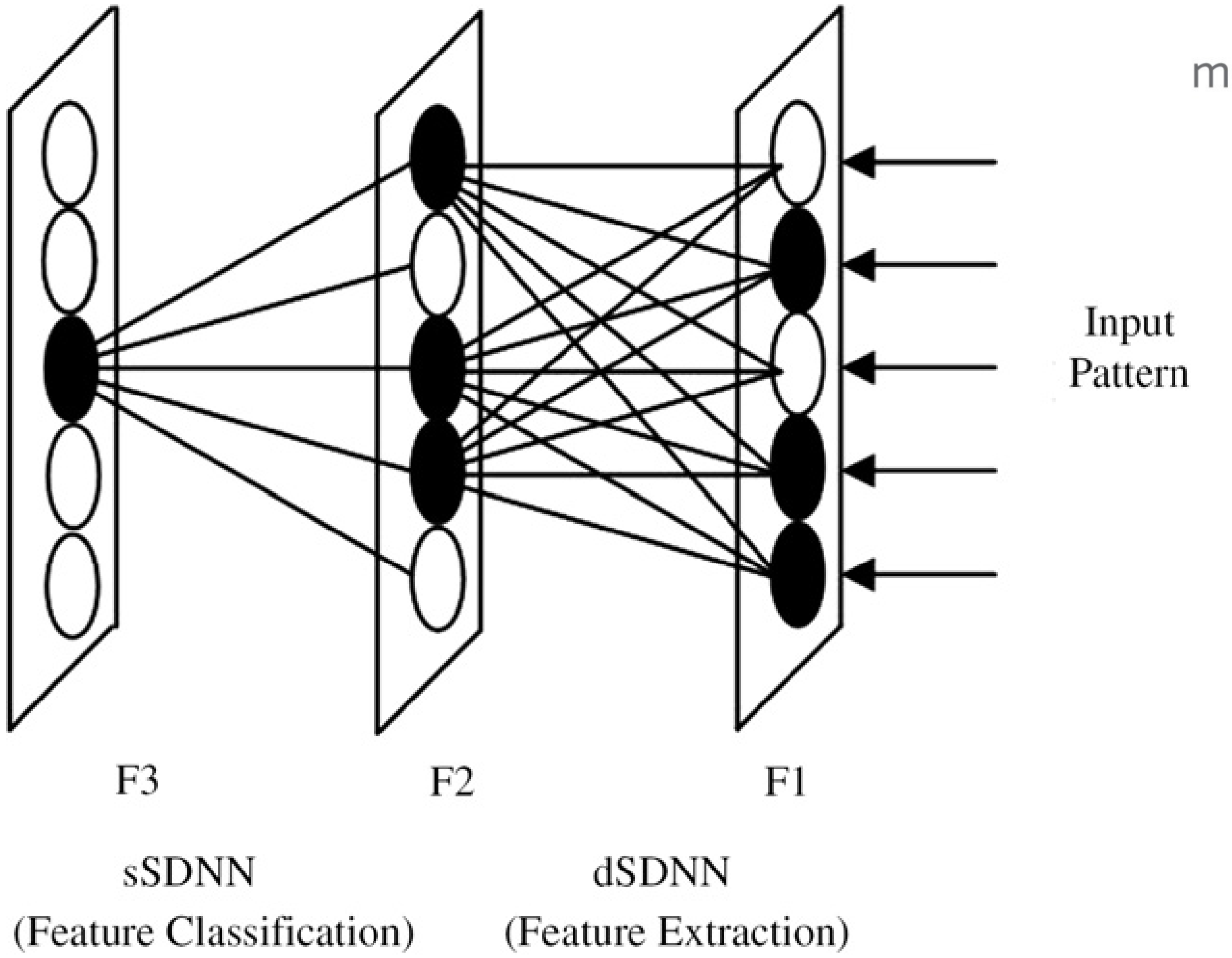


Fig. 2 Snap-Drift Neural Network (after [2])

Learning Process in Snap-Drift Neural Networks

- The Neural Network is structured in several layers which perform alternating computational tasks (quick “snap” and subsequent “drifts”)
- The first layer learns to group the input data according their feature, thus performing feature extraction. The second layer compares the input to a given threshold, thus finalizing the feature classification, or passes the input for further analysis of the pattern
- As a result, the NN can rapidly adapt to new data patterns. This makes the Snap-Drift NNs very suitable for online Web analytics. In addition they allow controlling the number of generated classes but require pre-formatting and normalization of the input data

3.3 Using Linguistic Thesaurus for Disambiguation of Search Terms

- Used for classification of the keywords searched on the Web sites, which relies on an online language *thesaurus* for semantic disambiguation of the words and phrases.
- Based on a version of the popular online thesaurus **WordNet**, developed by **George Miller** and his co-workers at Princeton University [3].
- Implemented experimentally as an online processing utility and originally planned for inclusion in **Decibel Insight** v. 2 but postponed due to the change in the Google policy concerning providing keywords statistics of its search engine.

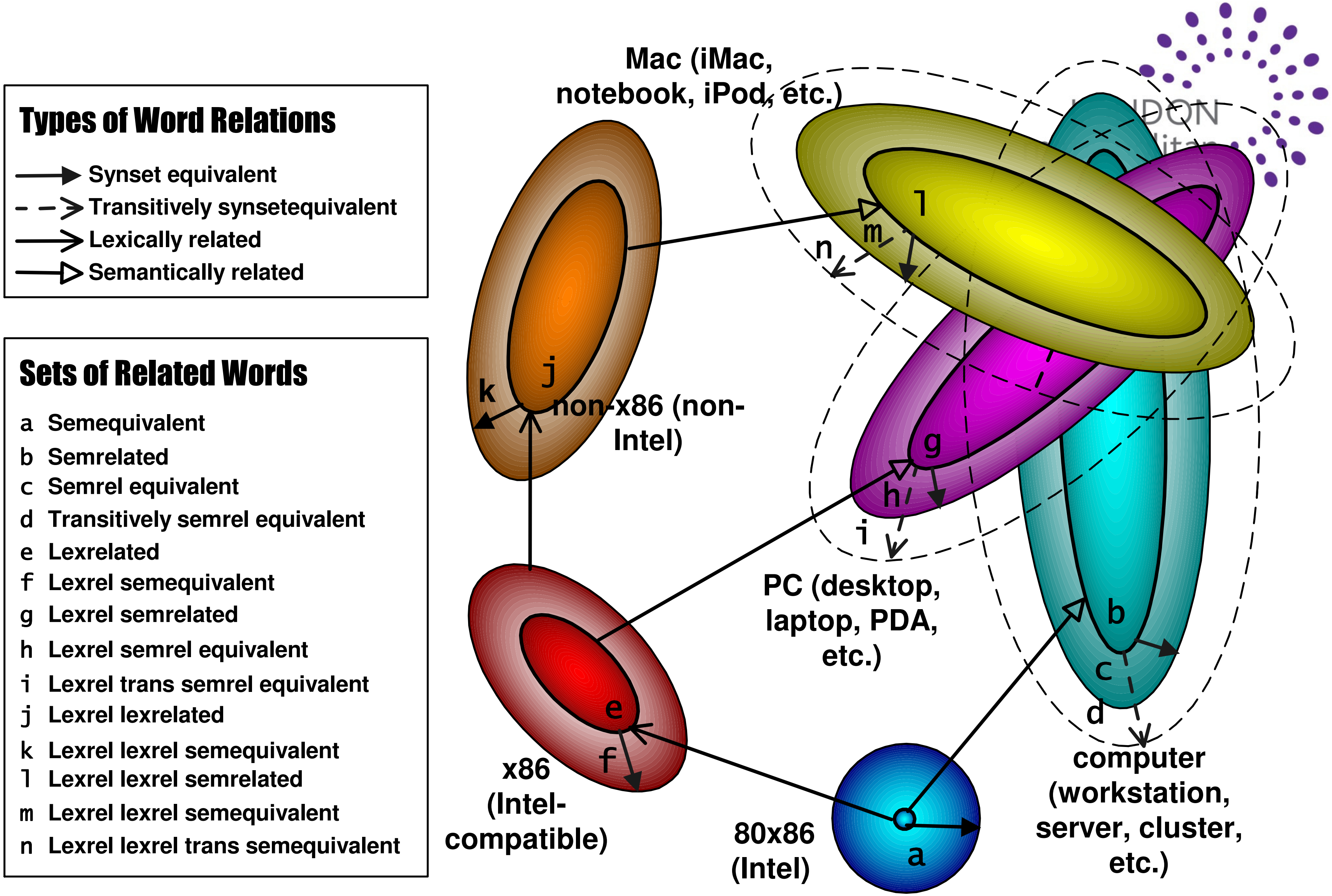
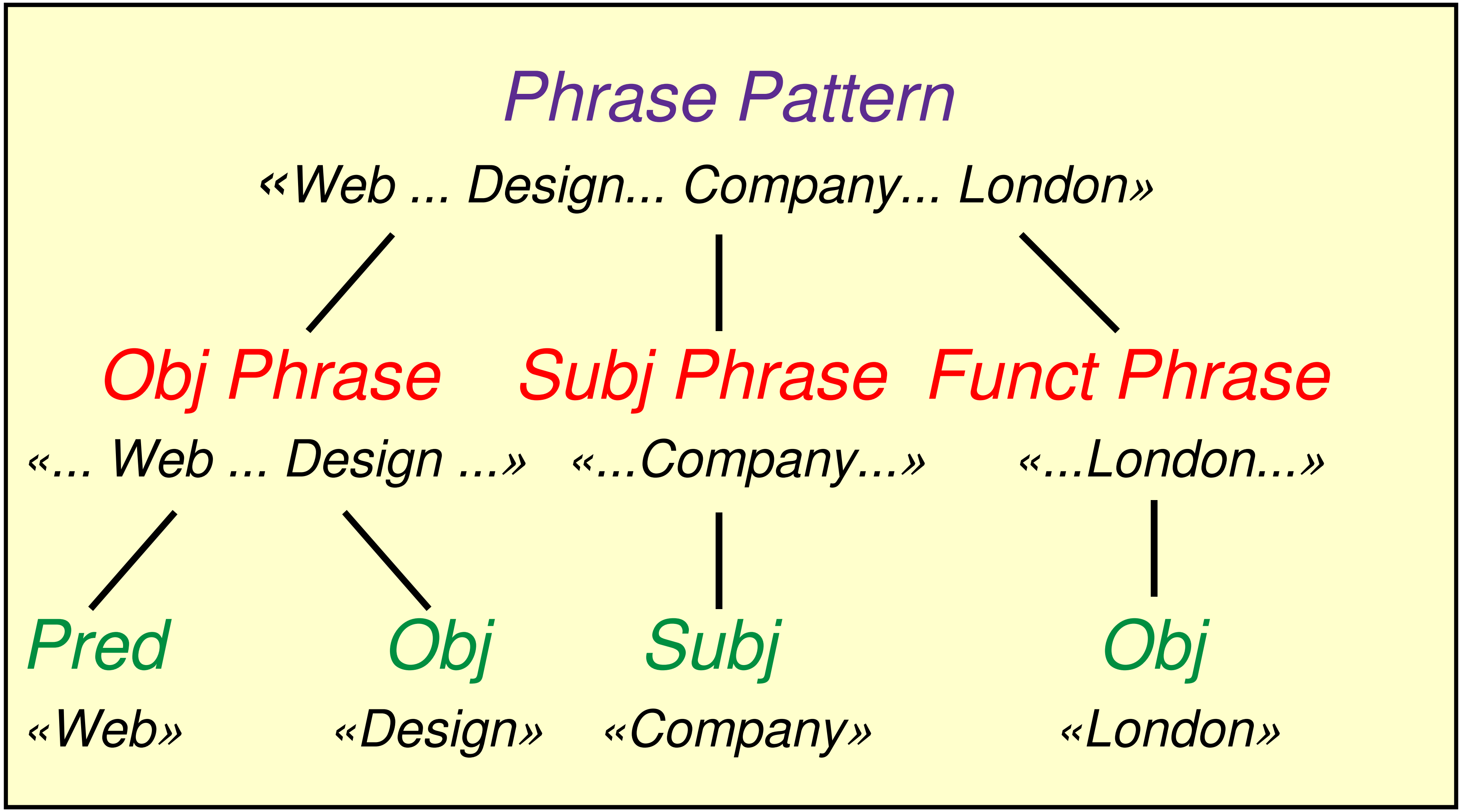

Types of Word Relations
Synset equivalent
Transitively synsetequivalent
Lexically related
Semantically related
Sets of Related Words
a Semequivalent
b Semrelated
c Semrel equivalent
d Transitively semrel equivalent
e Lexrelated
f Lexrel semequivalent
g Lexrel semrelated
h Lexrel semrel equivalent
i Lexrel trans semrel equivalent
j Lexrel lexrelated
k Lexrel lexrel semequivalent
l Lexrel lexrel semrelated
m Lexrel lexrel semequivalent
n Lexrel lexrel trans semequivalent
Mac (iMac, notebook, iPod, etc.)
l
m
n
k
j
non-x86 (non-Intel)
g
h
i
PC (desktop, laptop, PDA, etc.)
b
c
d
computer (workstation, server, cluster, etc.)
e
f
x86 (Intel-compatible)
80x86 (Intel)
a

Fig. 3 Word Relations in WordNet Thesaurus

Phrase Grammar for Keyword Classification



Phrase Pattern

«Web ... Design... Company... London»

Obj Phrase

«... Web ... Design ...»

Subj Phrase

«...Company...»

Funct Phrase

«...London...»

Pred

«Web»

Obj

«Design»

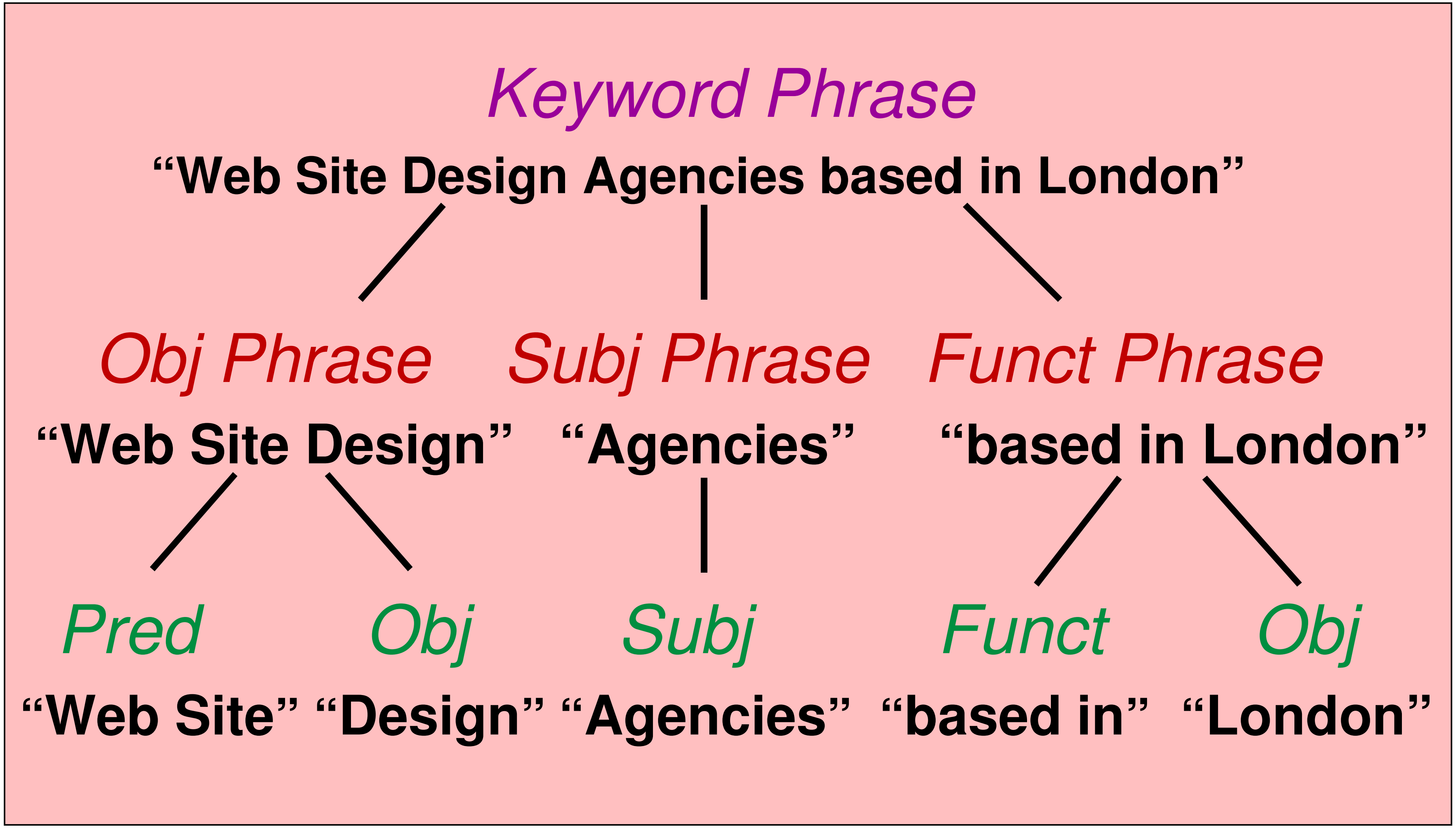
Subj

«Company»

Obj

«London»

An Example of a Search Keywords which meet the phrase Grammar



Keyword Phrase

“Web Site Design Agencies based in London”

Obj Phrase *Subj Phrase* *Funct Phrase*

“Web Site Design” **“Agencies”** **“based in London”**

Pred *Obj* *Subj* *Funct* *Obj*

“Web Site” **“Design”** **“Agencies”** **“based in”** **“London”**

4 Further Development

- Incorporating online data segmentation using the Snap-Drift NN. The classification engine is working, some work needs to be done to interpret the generated classes in market terms.
- Using No-SQL storage for incremental reconstruction of the journey paths and generating 3D maps of the user journeys based on the navigation paths. One experimental prototype has been developed using **Orient DB** and another one based on **Mongo DB** is on the way.
- Adding local indexing and search engine with internal control of the keyword search for further analysis of the user intentions. Some work has been done on the keyword search history using AJAX technology.

References

[1] IBM. IBM Tealeaf CX solutions [available online at http://public.dhe.ibm.com/common/ssi/ecm/zz/en/zzw03191usen/ZZW03191USEN.PDF]

[2] S.W. Lee, D. Palmer-Brown, C.M. Roadknight, Performance-guided neural network for rapidly self-organising active network management (Invited Paper), Journal of Neurocomputing 61C (2004), pp. 5–20.

[3] G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. WordNet: An online lexical database. Int. J. Lexicography 3/4 (1990), pp. 235–244.

Maximise your website.