

Information mining at FUB

Claudio Carpineto



Fondazione Ugo Bordoni



Information Mining area at FUB: who we are

Researchers

Gianni Amati

Andrea Bernardini

Claudio Carpineto (head)

Raffaele Nicolussi

Gianni Romano

PhD students

Giuseppe Amodeo

Carla Michini

Scholarship holders

Edgardo Ambrosi

Massimiliano D'Amico

Giovanni Stilo

Master thesis student

Raffaella Toscano





Main scientific achievements

New probabilistic ranking methods

DEVIATION
FROM
RANDOMNESS

*Excellent results at
TREC and CLEF
since 1999*

INFORMATION
THEORETIC
QUERY
EXPANSION

Information retrieval and mining with CONCEPT LATTICES


*Concept Data Analysis book
(2004)*



Developments of Terrier, an open source search engine

(in collaboration with University of Glasgow and Matrixware)

TERabyte RetrIEveR



Terrier

[Home](#) [About](#) [Features](#) [Download](#) [Documentation](#) [Forum](#) [Issue Tracker](#) [Publications](#) [People](#) [News](#)

Links

- [How to Contribute ?](#)
- [Terrier Team Blog](#)
- [Terrier Wiki](#)
- [Labrador](#)
- [Smooth](#)
- [References](#)

Search

Powered by **Terrier** and **Labrador**

Project Information

- ECIR'08 Tutorial
- Workshop Tutorial 05
- EPSRC Final Report
- Terrier Overview
- Terrier Poster
- Flyer Nov04
- Flyer Mar05
- Flyer Aug07


Terrier

Terrier is a highly flexible, efficient, effective, and robust search engine, readily deployable on large-scale collections of documents. Terrier implements state-of-the-art indexing and retrieval functionalities. Terrier provides an ideal platform for the rapid development of large-scale retrieval applications.

The open source version of Terrier provides a comprehensive, flexible, robust, and transparent platform for research and experimentation in text retrieval. The research put into Terrier constantly expands towards new branches of the wider information retrieval field, making Terrier an ideal, strong, modular, and state-of-the-art platform for developing, assessing, and evaluating new concepts and ideas.


Terrier is written in Java, and was used for Web and Enterprise search, Desktop, Intranet and Vertical search engines, as well as developing and evaluating novel large-scale text information retrieval techniques and applications.

Terrier is being developed in the [Department of Computing Science](#), at the [University of Glasgow](#).



University of Glasgow

Department of Computing Science



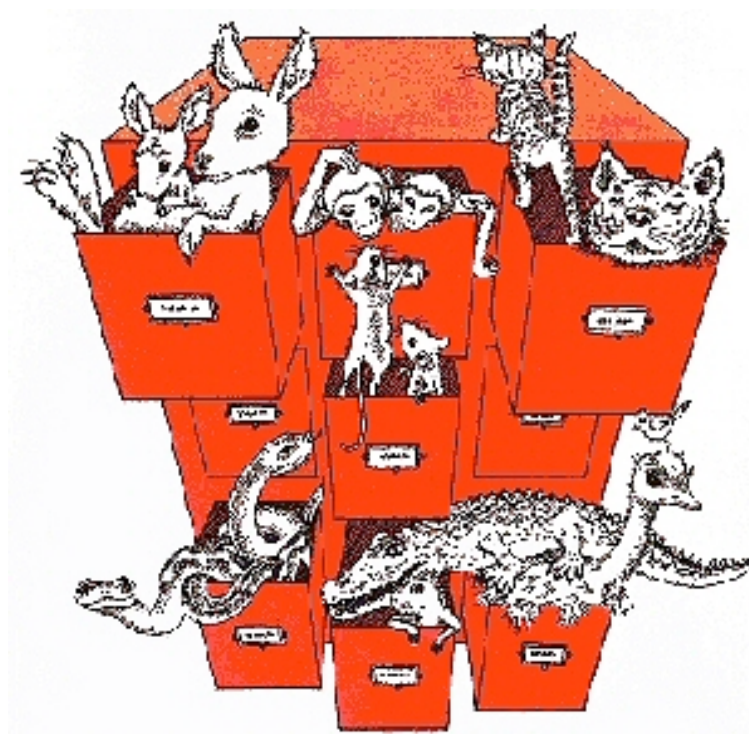
bug and issue tracker





Intelligent web search with


KeySRC





Keyphrase-based Search Results Clustering (1)

<http://keysrc.fub.it>



Home Preferences Links Documents Contact

data mining Search

All results (99)

- » data mining knowledge discovery (21)
- » data mining and predictive modeling (10)
- » data mining tools (6)
- » business intelligence (8)
- » data mining applications (7)
- » data mining techniques (5)
- » machine learning (5)
- » data mining helps (3)
- » mining software (8)
- » data mining solutions (5)

[More clusters](#)

TOP 99 RESULTS OF RETRIEVED FOR THE QUERY DATA MINING

DATA MINING - WIKIPEDIA

Article about knowledge-discovery in databases (KDD), the practice of automatically searching large stores of **data** for patterns.

http://en.wikipedia.org/wiki/Data_mining

DATA MINING: DEFINITION FROM ANSWERS.COM

data mining n. The automatic extraction of useful, often previously unknown information from large databases or **data** ... **Data Mining** For Investing ...

<http://www.answers.com/topic/data-mining>

DATA MINING - WIKIPEDIA, THE FREE ENCYCLOPEDIA

Data mining will not uncover patterns that are present in the domain, but not in ... The term **data mining** is often used to apply to the two separate processes of ...

http://en.wikipedia.org/wiki/Subject-based_data_mining

AN INTRODUCTION TO DATA MINING

Data mining, the extraction of hidden predictive information from large ... **Data mining** tools predict future trends and behaviors, allowing businesses to ...

<http://www.theartling.com/text/dmwhite/dmwhite.htm>

ELECTRONIC STATISTICS TEXTBOOK: DATA MINING TECHNIQUES

Outlines the crucial concepts in **data mining**, defines the **data** warehousing process, and offers examples of computational and graphical exploratory **data** analysis ...

<http://www.statsoft.com/textbook/stdatmin.html>

DATA MINING: TEXT MINING, VISUALIZATION AND SOCIAL MEDIA

Commentary on text **mining**, **data mining**, social media and **data** visualization. ... relevant publications in top **data mining**, information retrieval and natural ...

<http://datamining.typepad.com/>

DATA MINING: WHAT IS DATA MINING?

Outlines what knowledge discovery, the process of analyzing **data** from different perspectives and summarizing it into useful information, can do and how it works.


<http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>





Keyphrase-based Search Results Clustering (2)

<http://keysrc.fub.it>

Home Preferences Links Documents Contact

All results (99)

- » [data mining knowledge discovery](#) (21)
- » [data mining and predictive modeling](#) (10)
- » [data mining tools](#) (6)
- » [business intelligence](#) (8)
- » [data mining applications](#) (7)
- » [data mining techniques](#) (5)
- » [machine learning](#) (5)
- » [data mining helps](#) (3)
- » [mining software](#) (8)
- » [data mining solutions](#) (5)

[More clusters](#)

YOU ARE IN "DATA MINING TOOLS" CLUSTER WITH 6 DOCUMENTS

AN INTRODUCTION TO DATA MINING
Data mining, the extraction of hidden predictive information from large ... Data mining tools predict future trends and behaviors, allowing businesses to ...
<http://www.theartling.com/text/dmwhite/dmwhite.htm>

TWO CROWS WHITE PAPER: "SCALABLE DATA MINING"
White paper on the reasons scalable data mining solutions are needed and the main methods of achieving scalability. ... Making Data Mining Tools Scalable ...
<http://www.twocrows.com/whitep.htm>

OPEN DIRECTORY - COMPUTERS: SOFTWARE: DATABASES: DATA MINING: TOOL VENDORS
Advanced Software Applications - Data mining, analysis, and decision support ... Provides neural network software for data mining and forecasting as well as ...
http://www.dmoz.org/Computers/Software/Databases/Data_Mining/Tool_Vendors/

DATA MINING - DATA MINING SOFTWARE (SOFTWARE) - THE DATA MINE WIKI
All currently available data mining software, tools and applications whose ... You can add Data Mining tools / software to this list. Don't know where to start? ...
<http://www.the-data-mine.com/bin/view/Software/DataMiningSoftware>

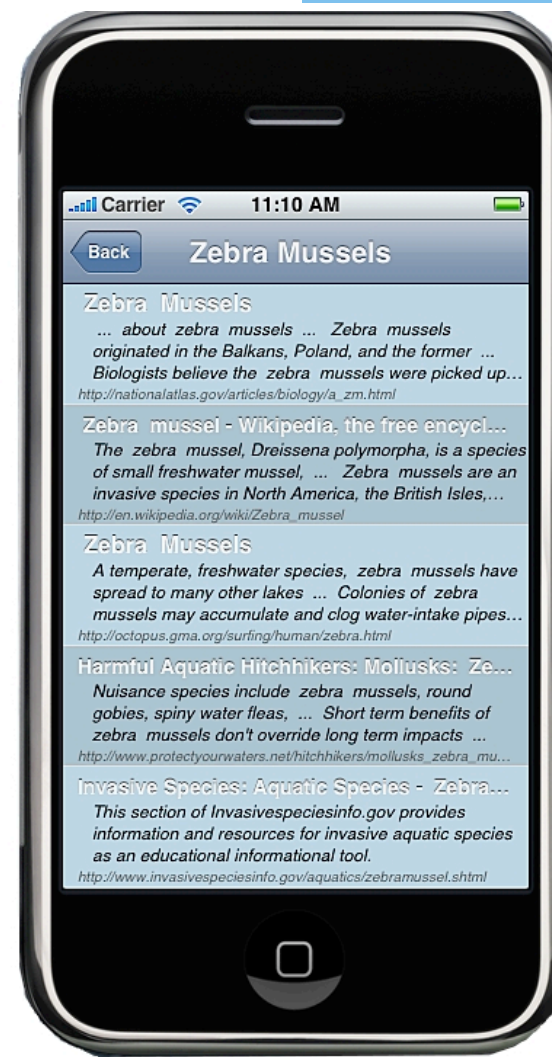
UNDERSTANDING DATA MINING
... theory, whereas data mining emerged from computational ... Data mining techniques do not require the user to know the ... Today's data mining tools ...
http://www.qbase.us/pdf/Technology/White%20Paper_Data%20Mining.pdf

A COMPARISON OF LEADING DATA MINING TOOLS
on Knowledge Discovery & Data Mining. Friday, August 28, 1998. New York, New York ... Compare and Summarize Data Mining Tools which: ...
http://www.datamininglab.com/pubs/kdd98_elder_abbott_nopics_bw.pdf





KeySRC on iPhone





Data mining for CRM (sales intelligence)

orangeTM

Server time : 2009-04-21 15:42

KDD Cup 2009

KDD-09
The 15th ACM SIGKDD Conference
On Knowledge Discovery
and Data Mining

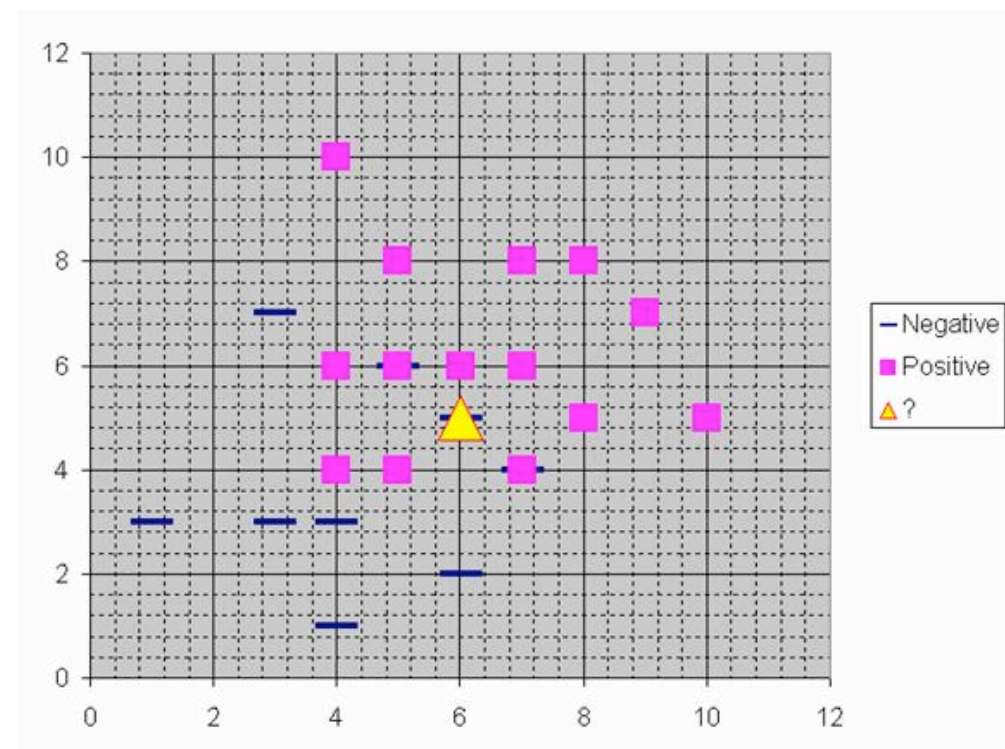
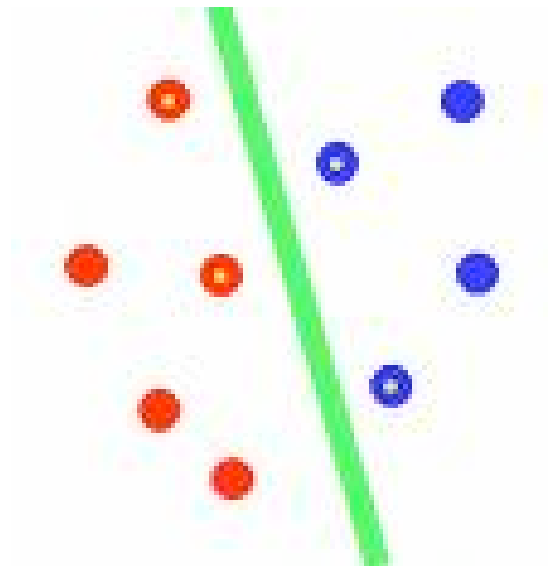
Index Data Instructions Forum Evaluation Submit Results Prizes Workshop

- *Churn*
- *Appetency*
- *Up-selling*
-





Data mining in theory





Data mining in practice



- High data *dimensionality* (millions of records, tens of thousands of attributes)
- *Missing* and *noisy* values
- Different value *scales*
- Different value *ranges* in training and test
- *Unbalanced* data set



Thank you





Thank you

