

# ZIMSKA ŠOLA FIŠ

# ODKRIVANJE ZNANJA V PODATKIH

Vsebina in pregled področja

Doc. dr. Matej Mertik



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# ZIMSKA ŠOLA FIŠ

## PREGLED PODROČJA ODKRIVANJA ZNANJA

## VSEBINA ŠOLE IN PREDAVATELJI



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# Zimska šola FIŠ: Odkrivanje znanja v podatkih

BODI IN!



## Kdo sodeluje?

Institut Jozef Stefan  
Univerza v Udinah, Italija,  
Univerza v Zagrebu Hrvaška,  
Univerza v Bologni, Italija,  
Podjetje RapidI



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

## Predavatelji:

**Doc. dr. Bernard Zenko, Institut Jožef Stefan, Odsek za tehnologije znanja**

**Dr. Martin Žnidaršič, Institut Jožef Stefan**

**Dr. Angelo Di Lorio, University of Bologna, Italija**

**Prof. dr. Maurizio Pighin, University of Udine, Italija**

**Prof. dr. Anna Marzonna, University of Udine, Italija**

**Dr. Ingo Mierswa, Podjetje Rapidl, Nemčija, CEO in strokovnjak poslovne analitike**

**Dr. Simon Fischer, Podjetje Rapidl, Nemčija, strokovnjak poslovne analitike**

**Dr. Markus Schatten, Univerza v Zagrebu, področje poslovne informatike**



Bernard Ženko



Angelo Di Lorio



Maurizio Pighin



Anna Marzonna



Markus Schatten



Ingo Mierswa



Simon Fischer



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# TEHNIČNA IZVEDBA

DVE SKUPINI

SKUPINA 1:

28.1.2013 – 1.2.2013

SKUPINA 2:

19, 21, 26, 28 februar 2013



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# DANI I



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# DAN 1:

## Odkrivanje znanja v podatkih

[ Bernard Zenko, IJS ]

**Jozef Stefan Institute**, Department of Knowledge Technologies.

**Reserch interest:** machine learning, rule learning, predictive clustering, ensemble methods.

*Introduction course to konwledge technologies and datamining*



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Koheznijski sklad  
Evropski socialni sklad

# DAN 2



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad



# DAN 2: Semantični splet

## [Angelo di Lorio, University of Bologna]

**Researcher** at the Department of Computer Science and Engineering (DISI) of the University of Bologna.

**Research interest:** markup languages and document engineering areas. pagination and sophisticated templating, simplified authoring and publishing processes, collaborative editing, versioning and diff-ing techniques, metadata and automatic generation of semantic annotations, web technologies.

**Visiting and research:** Scholarly Technology Group of the Brown University (Providence, RI, USA), HPLabs, Bristol, UK, XSL-FO Working Group of the W3C (World Wide Web Consortium). He contributed to the transition of the group into the XML Print and Page Layout group, co-chair of the 1st ACM Workshop on New Forms of Xanalogical Storage and Functions, co-chair of the Web Technologies track at the ACM Symposium on Applied Computing in the last three editions, member in several international conferences and workshops, among which ICWE, UMAP, SAC, ITA, HSI, HT.

# SEMANTIČNI SPLET

Web 1.0 (the “Information Web”) was characterized by:

- The development of **HTTP** and **HTML** by Sir Tim Berners-Lee in 1990
- E-mail
- **Web directories** and search engines (notably Yahoo and Google)
- The rapid proliferation of websites
- **Many readers**, but relatively few authors or content creators.



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# SEMANTIČNI SPLET

Web 2.0 (the “Social Web”) is characterized by:

- Privately owned social networking websites such as **Facebook** (launched in 2004) and Twitter (launched in 2006)
- **Web services**, which allow developers to combine (or “mash-up”) content from various websites (ex. Historypin: a site where users can upload photos from the past and pin them on Google Maps)
- **Cloud computing which** virtualizes computing resources making them affordable and ubiquitous
- **Browser-based applications**, which eliminate the need for managing desktop applications
- **Web syndication**, which enables content to be subscribed to and reused by having it identified, collected and combined (“aggregated”) using RSS, and more recently the Atom publishing protocol.



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# SEMANTIČNI SPLET

Web 3.0 (the “Semantic Web”) is emerging, and is characterized by:

- **Linked data or hyperdata**, where data objects are linked to other data objects (similar to how web pages are linked today)
- **Large hyperdata datasets** such as DBpedia (a community effort to extract structured information from Wikipedia and make the information available on the Web)
- **A query language** for hyperdata capable of treating the entire web as a single datastore, called SPARQL
- The so-called “**Internet of Things**” where billions of non-human entities (including houses, cars and appliances) generate and publish their own hyperdata.



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# PREGLED VSEBIN

## [Semantični splet]

### DEL I

- glavni koncepti semantičnega spleta
- zakaj semantični splet, motivacija
- standardi za konceptom semantičnega spleta  
W3C
  - RDF, OWL in SPARQL
- raziskovanje in integracija heterogenih zbirk znanja
- koncept Linked Open Data



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# PREGLED VSEBIN

## [Semantični splet]

### DEL II

- Tehnologije za zajem sematičnega spleta
- Formati RDF, Microformat, Microdata
- Semantični wiki
- Web Commons Data, razvoj v prihodnje za potrebe ekstrakcije informacij in znanja iz spleta



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# DAN 3



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# DAN 3: Pridobivanje informacij iz procesov in podatkov v poslovnem svetu

[Mauritzio Pighin, Anna Morzonna, University of Udine]

**Professors** at University of Udine in Department of Mathematics and Computer Science.

**Reserach interest:** areas of bussnies informatics, knowledge technologies and datamining, reserch in unlearning and relearning in processes of business information systems and innovation.

**Courses:** area of bussnines information systems, innovation and datamining processes.

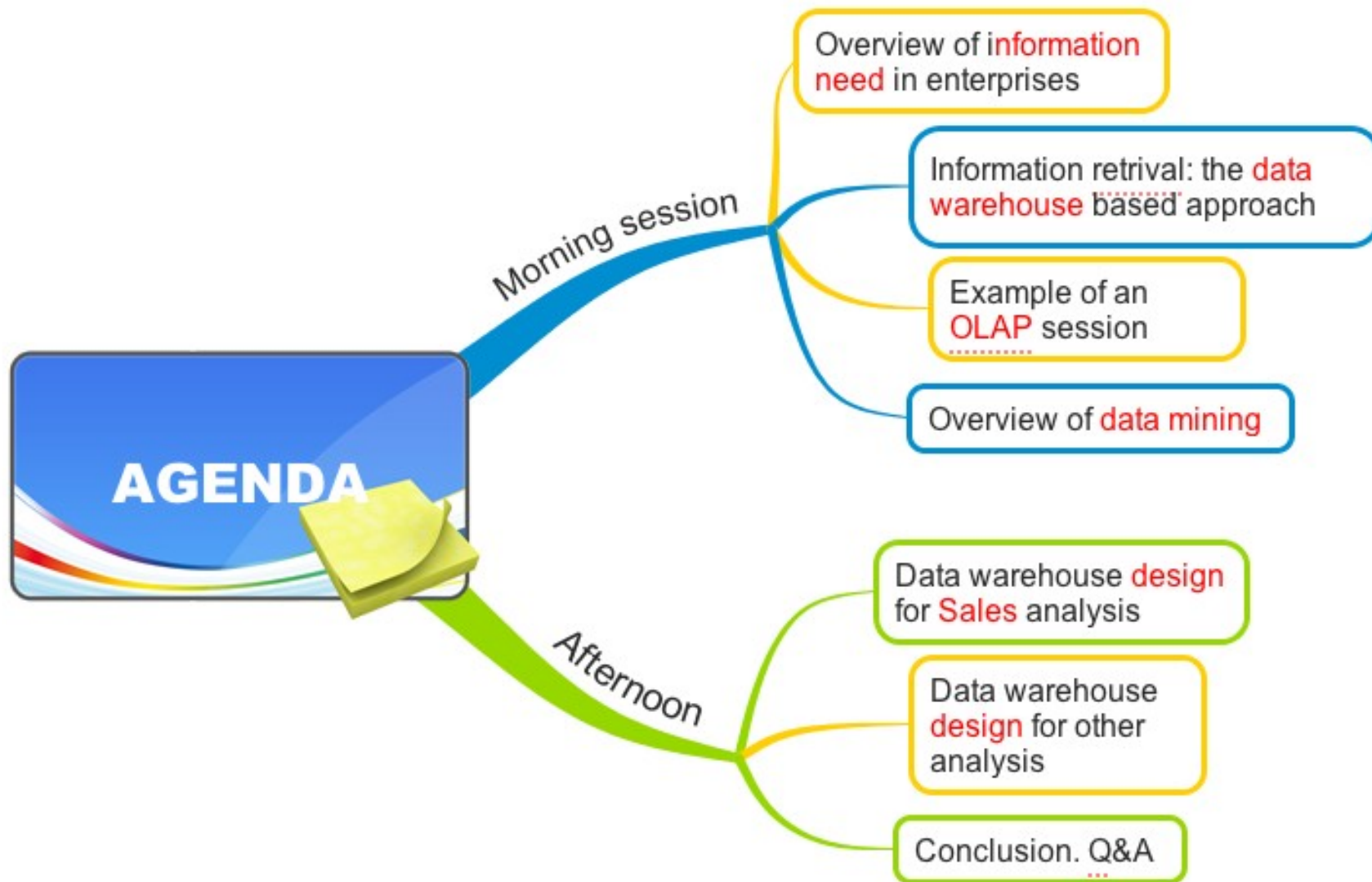


*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad



# Overview of lecture



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# Business information systems

- Enterprise structure
- Information need in enterprise
- Operational data sources
- Processes and data
  - Sales
  - Procurement
  - Logistics
  - Production
  - Accounting
- Flows and interactions



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# Data Warehousing and elements of Data Mining

- What is a data warehouse?
- A multi-dimensional data model
- Data warehouse architecture
- OLAP analysis
- From data warehousing to data mining
- Principles of data mining



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# Business processes and business data in companies

- Sales
- Procurement
- Logistics
- Production
- Accounting

DAN 4

DAN 5



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad

# DAN 4 in 5: Primeri in delo z orodjem RapidMiner

[Simon Fischer, Ingo Mierswa, Rapid I, Matej Mertik FIS]

**Ingo Mierswa** is the founder of Rapid-I and CEO of Rapid-I Germany. He has several years of experience in project management, human resources management, consulting, and leadership including ten years of coordinating and leading the multi-national RapidMiner developer team with about 50 developers and contributors world-wide.

**Simon Fischer** is Head of Research and Development at Rapid-I. He is one of the initiators of the RapidMiner project started in 2001 which has by now, together with the data mining server RapidAnalytics, become the number one open source data mining solution. Simon Fischer leads the development of the Rapid-I product suite and coordinates various research projects at Rapid-I.



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Koheznijski sklad  
Evropski socialni sklad

# RapidMiner

HOME CONTACT US SEARCH SITEMAP PRIVACY POLICY IMPRINT 



**The No 1 for Analysis**  
Try our Data Analysis Solution  
RapidMiner - Learn More!



PRODUCTS DOWNLOADS DOCUMENTATION SERVICES COMMUNITY PARTNER SHOP ABOUT US

## Products

- [Product Overview](#)
- [RapidMiner](#)
- [RapidAnalytics](#)
- [RapidNet](#)
- [RapidSentilyzer / BuzzBoard](#)
- [RapidDoc](#)
- [Extensions](#)



RapidMiner is unquestionably the world-leading open-source system for data mining. It is available as a stand-alone application for data analysis and as a data mining engine for the integration into own products. Thousands of applications of RapidMiner in more than 40 countries give their users a competitive edge.



 Všeč mi je  Osebam Olivera Grjevic, Ingo Mierswa in 2.738 drugim je to všeč.

## Testimonials

"I just found your RapidMiner software"

### Overview

### Features

### Screenshots

### Testimonials

### Editions and Prices

- **Data Integration, Analytical ETL, Data Analysis, and Reporting** in one single suite
- Powerful but intuitive **graphical user interface** for the design of analysis processes
- **Repositories** for process, data and meta data handling



*Naložba v vašo prihodnost*

OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj  
Kohezijski sklad  
Evropski socialni sklad