

# Applied informatics

Acquisition and exploitation of  
information from freely available  
database sources in a selected  
professional topic.

ZEMÁNEK, Z. – PLUSKAL, D. – SMETANA, B.

# **Acquisition and exploitation of information from freely available database sources in a selected professional topic.**

- 1. Selection and restriction of information**
- 2. Specification of a professional topic**
- 3. Options for database resources for professional topic**
- 4. Assignments**

# Aims of the lecture



1. Provide students with knowledge of the selection and restriction of information.
2. Introduce and explain the basis of the specifications of professional topics.
3. Clarify options for database resources for professional topics.

# Working with information

- ❑ Each set of real objects and phenomena has its rules, its placement in the hierarchy of the world, their classification into subtypes, their relationship to the environment.
- ❑ Also, a subset of attributes can have important relationships between them - association: correlation, cause and effect, hidden factors, etc., which are called variables in mathematical disciplines.
- ❑ A description of these attributes is often quite vague.
- ❑ For example: To describe a man about whom we speak about, we can say: "... the young blond man with shaggy bangs, probably from a neighboring door."

# Working with information

- ❑ Once we begin to explore some of the deeper phenomena systematically we begin by observing reality, collecting data, and an investigation of the facts on which the data is valid.
- ❑ We verify whether the facts can formulate generally applicable rules, or even do some (natural, social, ...) patterns.
- ❑ We gain the ability to create information and to pass on these.

[1]

# Selection, aggregation and restriction of information

- ❑ In the communication process there is always a selection of information that is usually affected by the needs and wishes of the recipient.
- ❑ At the same time this leads to accumulation (aggregation) of data into groups or aggregates.
- ❑ Thus limiting measures forced mostly by time-consuming data processing and limited storage capacity can lead to further, imprecise information processing.

# Selection, aggregation and restriction of information

- ❑ The next step is the restriction – this means limitations, such as data reduction of image conversion, and similar limiting measures. It is used for the constraints imposed by own or other authority's decision, not the real constraints and shortcomings.
- ❑ The result of the previous phenomena is that the information can often be distorted.
- ❑ The interpretation of the data follows, which consists of analysis of the importance of communication to the recipient. [2,3]

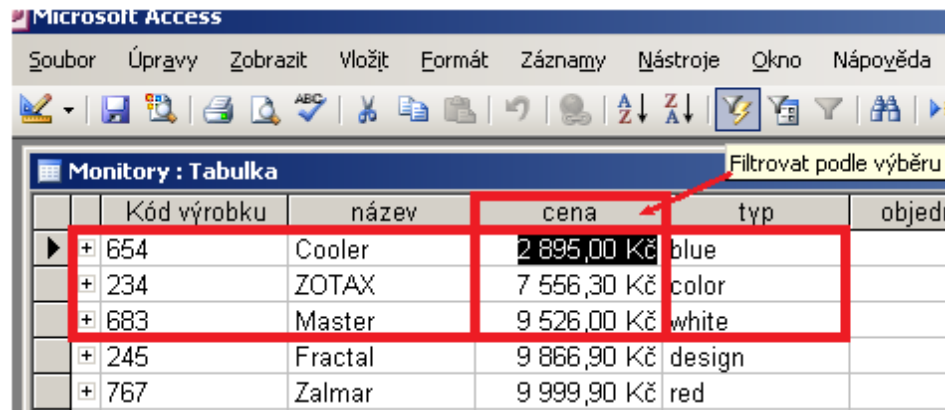
# Database access

- ☐ We work with an ordered data type.
- ☐ The initial selection process is simple.
- ☐ The main feature is the narrowing of records to be displayed according to the information of interest.
- ☐ Information interest is formulated as an information query.
- ☐ For a simple selection tools can be used, referred to as filtering, associated with the aggregation of data - displays the records arranged according to selected characteristics.



# Simple selection

- ❑ Select the first **n** records sorted by the property **x**
- ❑ Example: Select first **n** (= 3) records according to the lowest price [4]



Microsoft Access

Soubor Úpravy Zobrazit Vložit Formát Záznamy Nástroje Okno Nápověda

Monitory : Tabulka Filtrovat podle výběru


	Kód výrobku	název	cena	typ	objedr
+	654	Cooler	2 895,00 Kč	blue	
+	234	ZOTAX	7 556,30 Kč	color	
+	683	Master	9 526,00 Kč	white	
+	245	Fractal	9 866,90 Kč	design	
+	767	Zalmar	9 999,90 Kč	red	

# General restriction

- ❑ We select records according to the overall importance for the information query – e.g. simple existence of an attribute [4]
- ❑ Example: Listing suppliers, excluding ALZA ...

A	B	C
	ALZA	

výrobce : Tabulka			
	Číslo výrobce	název	adresa
	0987	CZC	Slaná 13
	6746	Bohemia	Jalovcová 5
▶	8634	ALZA	...
	8795	T-r	
	9567	Vd	

 Filtrovat podle výběru  
Filtr mimo výběr

# Restriction & projection

- ❑ We restrict the number of displayed attributes – projection.
- ❑ Example: selecting the product code without displaying the price and the supplier [4]

A	B	C

Monitory Dotaz : Výběrový dotaz			
	Kód výrobku	název	cena
▶	234	ZOTAX	7 556,30 Kč
	245	Fractal	9 866,90 Kč
	654	Cooler	7 835,80 Kč
	683	Master	9 526,00 Kč
	767	Zalmar	9 999,90 Kč

# Selection tools

- ❑ To create a selection or restriction we use an internal tool - queries.
- ❑ Queries using query language, which has its own syntax.
- ❑ The basis of the simple commands for selecting and operators "and", "or" not ...
- ❑ Using propositional logic we can assemble any meaningful statement – a query.

# Why query language?

- ❑ Natural language expresses emotions, and the degree of uncertainty is reduced by exploitation of non-verbal communication, professional speech, argot ...
- ❑ For machines this is inappropriate.
- ❑ The statement - sentence, the basis of the selection query.
- ❑ Complex sentences - the atomic decomposition, simple requirements.

❑ Atomic requirements for using operators than

# Query language

- ❑ It is a tool to process requests to find records that meet certain formally defined criteria.
- ❑ At the same time the instructions can be to calculate certain values on the basis of this matching records.

# Example of query language

**SQL** - Structured Query Language

❑ Retrieving data:

SELECT attributes FROM table WHERE conditions;

# Example of query language

## **QBE** Query By Example

- ❑ uses forms;
- ❑ query is defined by filling in a form having the database structure [6]

Family n.	First n.	ID	Math	English	Exam
»	»	XXXXXXXX	>=80	>=80	1 - 4



# Selection in full text

- ❑ The problem of selecting the data is in the irregular appearance in the text.
- ❑ Selection and restriction in fulltext:
  - ❑ Proper keyword selection and lemmatization for simple selection.
  - ❑ Use of languages and tools for the use of operators, use a searchable text in XML combined with restrictions.

# Problems with fulltext

- ❑ Creating the right lemma
- ❑ = Lemma (linguistics) - the basic form of the word or phrase
- ❑ Psychological barrier - Concentration on the homepage
- ❑ The language barrier in finding sources Translation and keywords
- ❑ The intentional manipulation of resources by authors - Keywords in the background, provide different text for the search robots ...

# Problems with fulltext

- ☐ Inadequate content of snippets.
- ☐ The problem with the concentration of resources in the invisible = not available Web.
- ☐ Unsuitable search environment in the database of text sources - bibliography ...
- ☐ Lack of knowledge about popular issues and semantics ...

# Search tools for fulltext

- ❑ One of the important tools is the search system – a product of the Company Tovek. It is used by the U.S. army.
- ❑ It uses indexing and ranking of documents.
- ❑ In addition to the operators and the possibility of a restrictive definition of subqueries it examines the relevance of the document to the query (weight). [7]

# How to search for professional information

## Motto:

„If you want to build a big company, build yourself first. “

Tomáš Baťa

- ☐ Collection of professional knowledge is a part of your project.

# Professional orientation

- ❑ It is an ambitious long-term process of preparation, choice and keeping a job.
- ❑ A part of the process, which is emphasized by the European Union, is testing your own assumptions, abilities and interests.
- ❑ A prerequisite is a good knowledge of professional themes and art to search, identify and analyze information from the professional sphere.

# Professional topics

- ☐ The first prerequisite for the success of professional orientation is the familiarity with the professional themes and terminology of the chosen profession.
- ☐ Very important is to work on your personal assumptions in the following areas:
  - ☐ work with a vision as motivation,
  - ☐ values and motivation as a prerequisite to the effectiveness of work
  - ☐ work under stress and in conflict situations,
  - ☐ communicative skills,
  - ☐ time management,
  - ☐ self-awareness, self-management and self-control.

[8]

# How to search for professional information

Finding professional information is concentrated in two areas:

Finding a suitable job within your chosen career - here you can also get an overview of the requirements: the required education, knowledge, skills and practice.

Advanced knowledge in the area of expertise:  
Terminology and professional/technical literature and open sources ... [8]

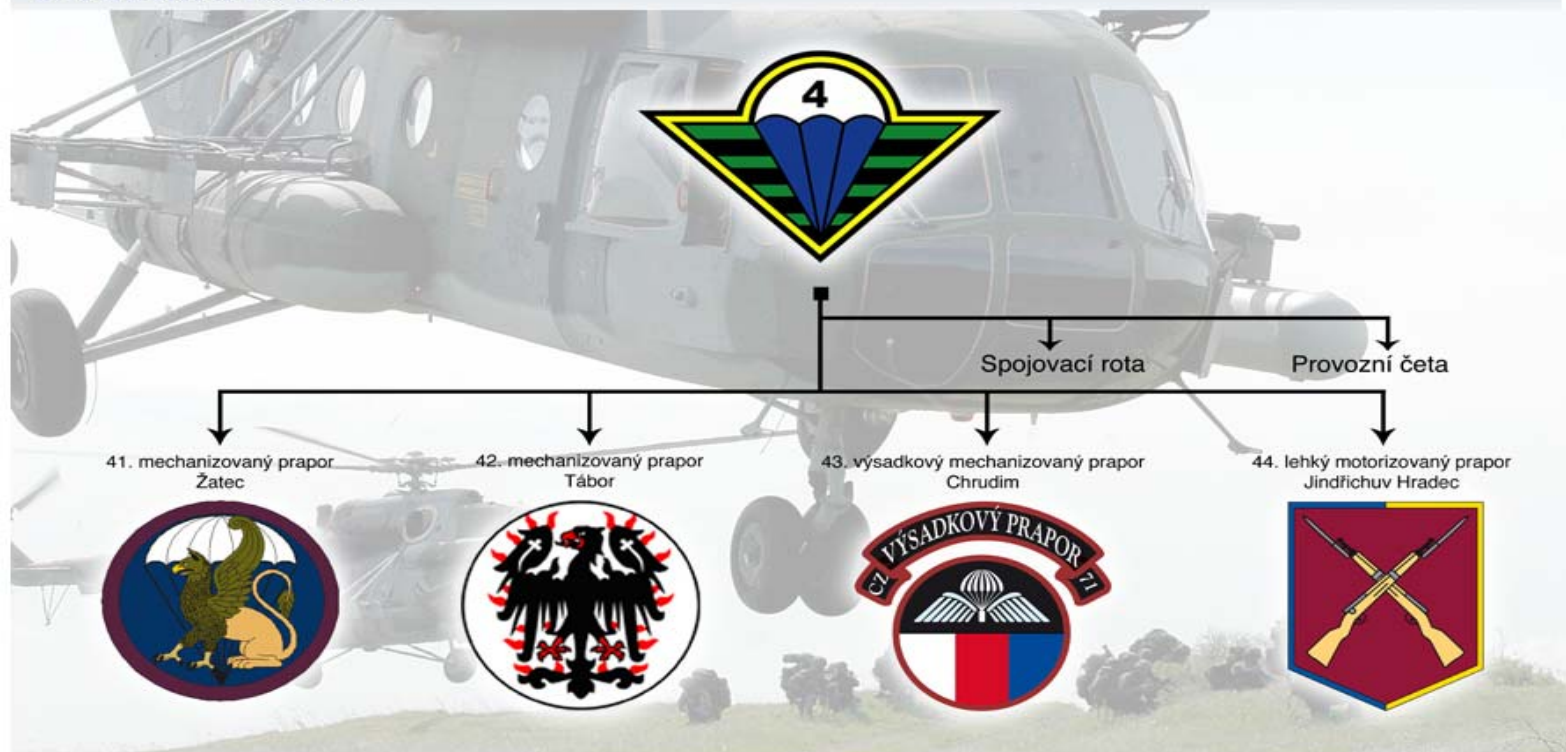


# Database sources

- ❑ Information can be found both in fulltext databases, such as database systems, sorting the menu according to different criteria in the form of advertisements – snippets - with the given structure.
- ❑ Possible environments are both official applications of MPSV, secondly agencies, ESF project and sector portals such as the [www.army.cz](http://www.army.cz).

# Example – [www.army.cz](http://www.army.cz)

## Struktura 4. brn



# Assignments



Analyze the possibilities of restrictions and selection of getting the professional information.



Be aware of the possibility of professional orientation in the chosen profession - specialization.



Create a list of basic professional concepts for the resource discovery.

# Resources:

1. ŠARMANOVÁ, J. *METODY ANALÝZY DAT - Učební text*. [online]. [cit. 2013-11-10] © 2012, Ostrava: VŠB-TU. 170 s. ISBN 978-80-248-2565-6 Dostupné z: <http://www.person.vsb.cz/archivcd/FEI/MAD/>
2. BARTÁK, M. *Vliv komunikace na konkurenceschopnost podniku pomocí nástrojů business intelligence – Disertační práce*. [online]. [cit. 2013-11-13] © 2006, Praha: ČZU-PEF. 185 s. ISBN 978-80-248-2565-6 Dostupné z: [www.pef.czu.cz/cs/?dl=1&f=13008](http://www.pef.czu.cz/cs/?dl=1&f=13008)
3. CATS-BARIL, W.-THOMPSON, R. *Information technology and Management*. USA: The McGraw-Hill Companies, Inc., ISBN 0-256-17618-3
4. Databáze standardu SQL, díl 8.: Restrikce v SQL. [Http://www.penguin.cz/](http://www.penguin.cz/) [online]. 1999, 2013 [cit. 2013-11-14]. Dostupné z: <http://www.penguin.cz/noviny/chip/sql/SQL8.pdf>
5. SQL. [Http://cs.wikipedia.org](http://cs.wikipedia.org) [online]. 2013 [cit. 2013-11-15]. Dostupné z: <http://cs.wikipedia.org/wiki/SQL>
6. QBE (Query By Example). [Http://www.cecak.cz](http://www.cecak.cz) [online]. 2009 [cit. 2013-11-15]. Dostupné z: [http://www.cecak.cz/fel/dba/referaty/qbe\\_query\\_by\\_example](http://www.cecak.cz/fel/dba/referaty/qbe_query_by_example)
7. Querylanguage\_cz. [Http://muj.anopress.cz](http://muj.anopress.cz) [online]. 2011 [cit. 2013-11-18]. Dostupné z: [http://muj.anopress.cz/Search/PagesHelp/querylanguage\\_cz.pdf](http://muj.anopress.cz/Search/PagesHelp/querylanguage_cz.pdf)
8. Profese-povolani. [Http://www.orientexpress.cz](http://www.orientexpress.cz) [online]. 2007 [cit. 2013-11-18]. Dostupné z: <http://www.orientexpress.cz/hledam-informace-z-oblasti/>