

Магистърска Програма:
“Извличане на информация и
откриване на знания”
(Information Retrieval and
Knowledge Discovery)

доц. д-р Иван Койчев
e-mail: [koichev \(AT\) fmi.uni-sofia.bg](mailto:koichev(AT)fmi.uni-sofia.bg)

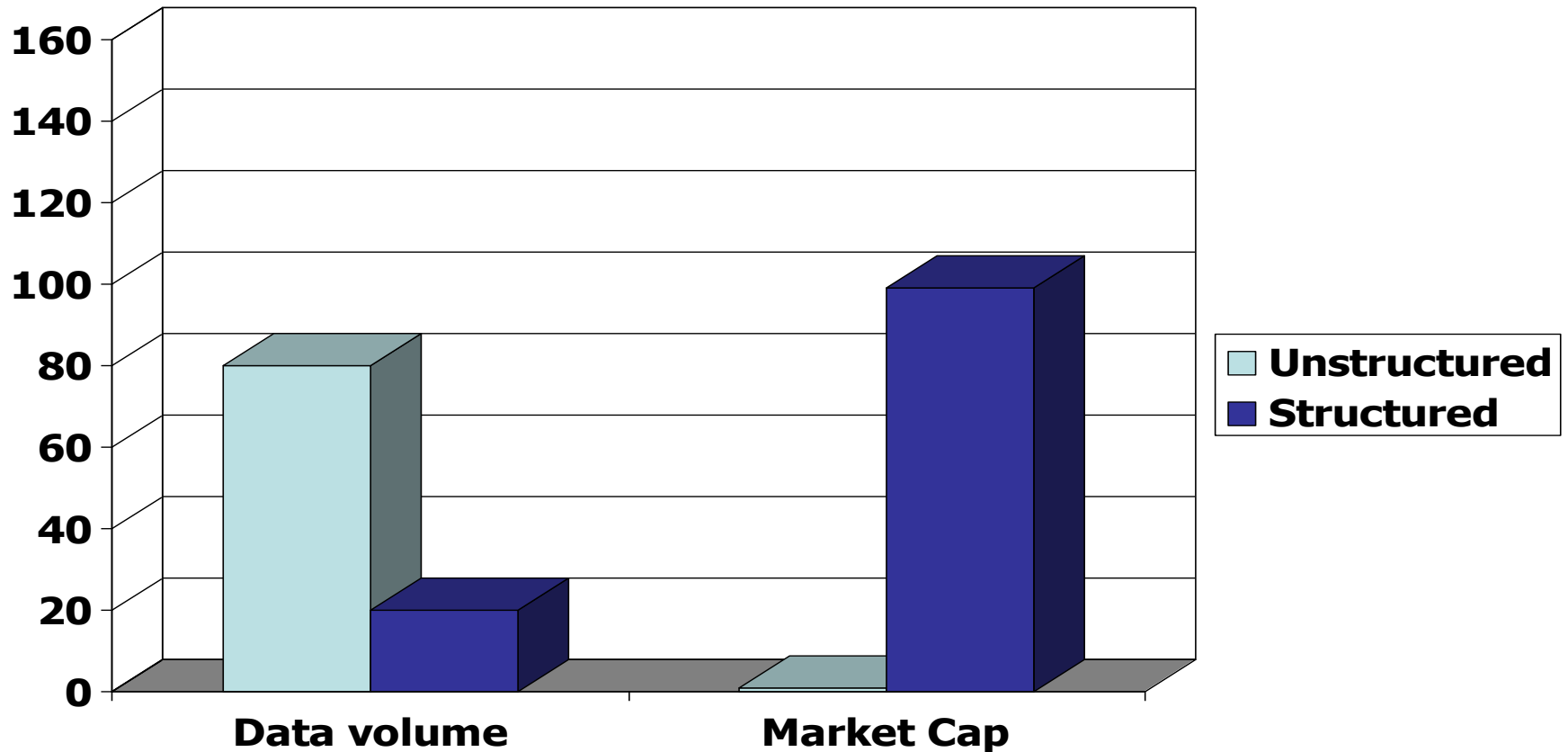
I. Извличане на Информация (Information Retrieval)

- Information Retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).

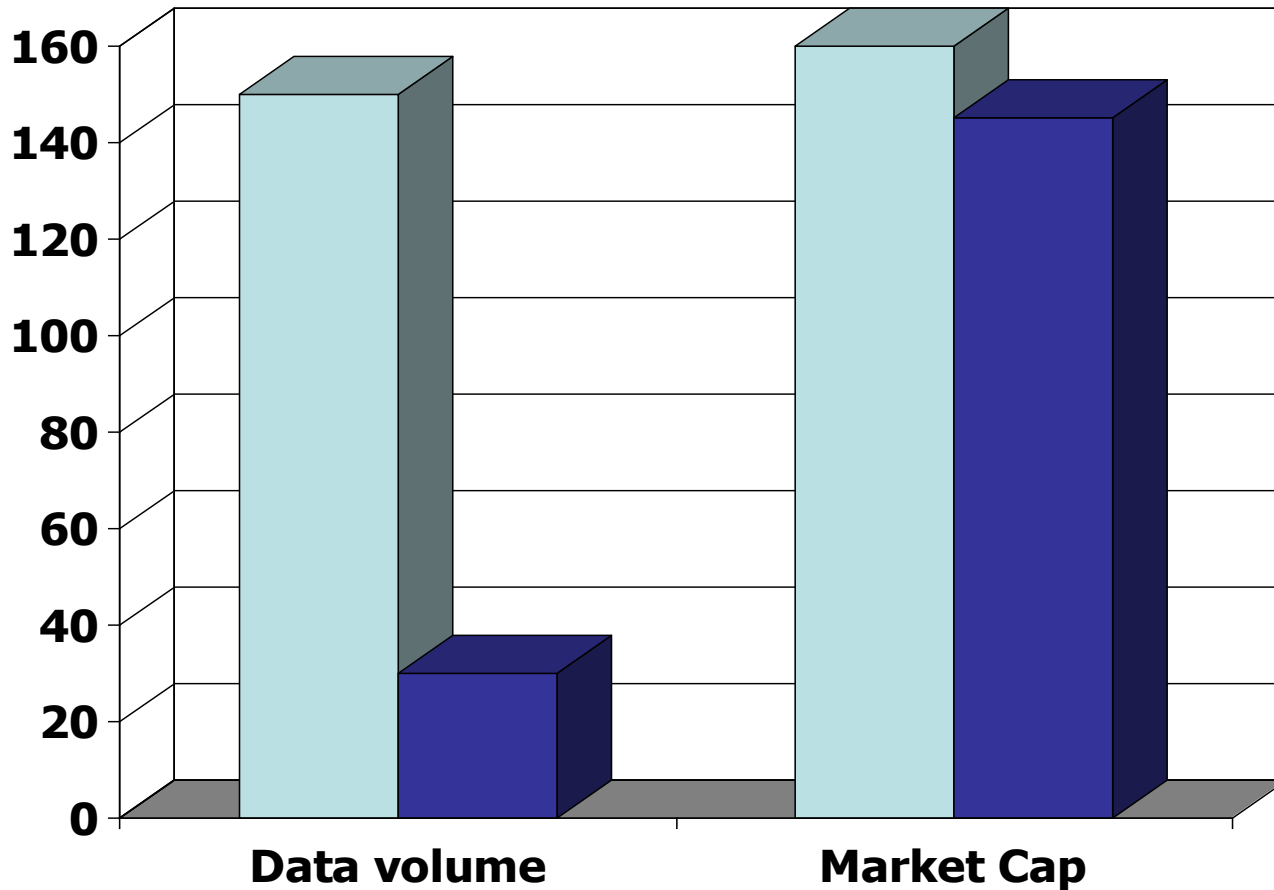
Защо да уча “Извличане на информация” ?

- Какво ще науча от този курс?
 - Как се правят системите за търсене и намиране на документи в океана от неструктурирана информация (текст)
- Има ли пазар за мен?
 - Вижте следващите 3 слайда и решете.

Unstructured (text) vs. structured (database) data in 1996



Unstructured (text) vs. structured (database) data in 2006



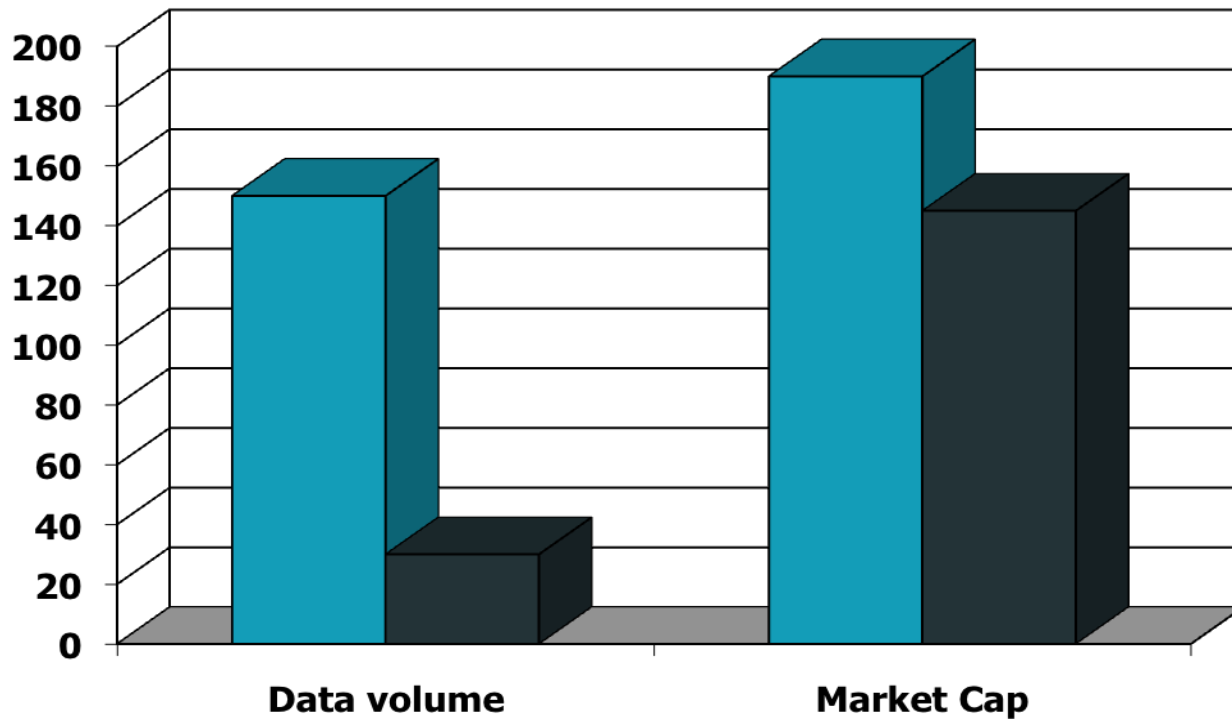
Google™

YAHOO!®

Unstructured
Structured



Unstructured (text) vs. structured (database) data in 2009



Google™

YAHOO!®

■ Unstructured
■ Structured

bing™

Ask™
.com

II. Откриване на знания

- Откриване на знания в данни
(Knowledge Discovery from Data, Data Mining, Analytics, ~ Business Intelligence, Predictive ...)
- Откриване на знания в текст
(Knowledge Discovery from text, Text Mining, Information Extraction, ...)
- Hot Topics: Web Mining, Opinion Mining, Sentiment Analysis, ...

Защо да уча “откриване на знания”?

- Обществото произвежда огромно количество данни (структурирани и неструктурирани)
 - Източници: бизнес, наука, медицина, икономика, география, околната среда, социален уеб, спорт, ...
- Потенциално ценен ресурс
- Суровите данни са **непотребни**: нужни са техники за автоматично извличане на информацията/знанията от тях.

Ще има ли пазар за мен?

- Списание Forbes 2/17/2012:
“Big Data Is Big Market & Big Business - **\$50 Billion Market by 2017**”:
 - “We all know that Big Data is the hottest sector in IT at the moment.”
 - “Big Data market at just over \$5 billion in total factory revenue.”
 - “Further, Wikibon forecast the Big Data market will grow at an astounding **CAGR of 58%** between now and 2017, hitting the \$50 billion within five years”

Shortage of Skills

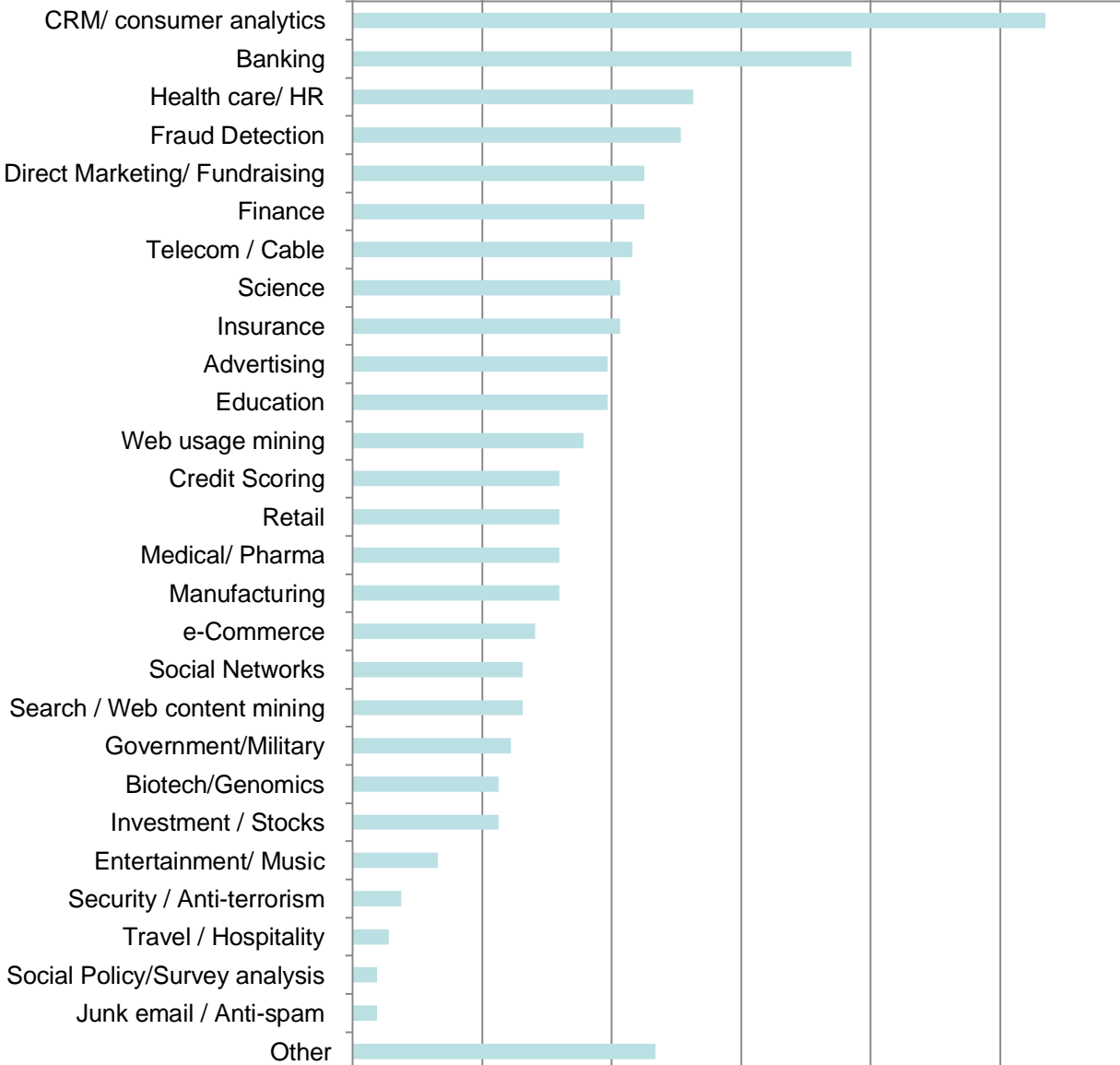
- McKinsey: shortage by 2018 in the US of
 - 140-190,000 people with deep analytical skills
 - 1.5 M managers/analysts with the know-how to use the analysis of big data to make effective decisions.

Source:

www.mckinsey.com/mgi/publications/big_data/

Where did you apply analytics/data mining?

0,0% 5,0% 10,0% 15,0% 20,0% 25,0% 30,0%



avg 2.4
industries

Задължителни курсове

- **Зимен семестър**
 - **Извличане на информация** (доц. И. Койчев и &)
 - **Подходи за обработка на естествен език**
 - (проф. Галя Ангелова и &)

- **Летен семестър**
 - **Откриване на знания в данни** (доц. И. Койчев и&)
 - **Откриване на знания в текст** (д-р Преслав Наков, д-р Георги Георгиев и &)

Избираеми курсове

- От областите на:
 - Изкуствен интелект
 - Софтуерни технологии
 - Статистика
 - ...

Благодаря Ви за вниманието

За информация и въпроси:

доц. Иван Койчев

e-mail: [koichev \(AT\) fmi.uni-sofia.bg](mailto:koichev(AT)fmi.uni-sofia.bg)