

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

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

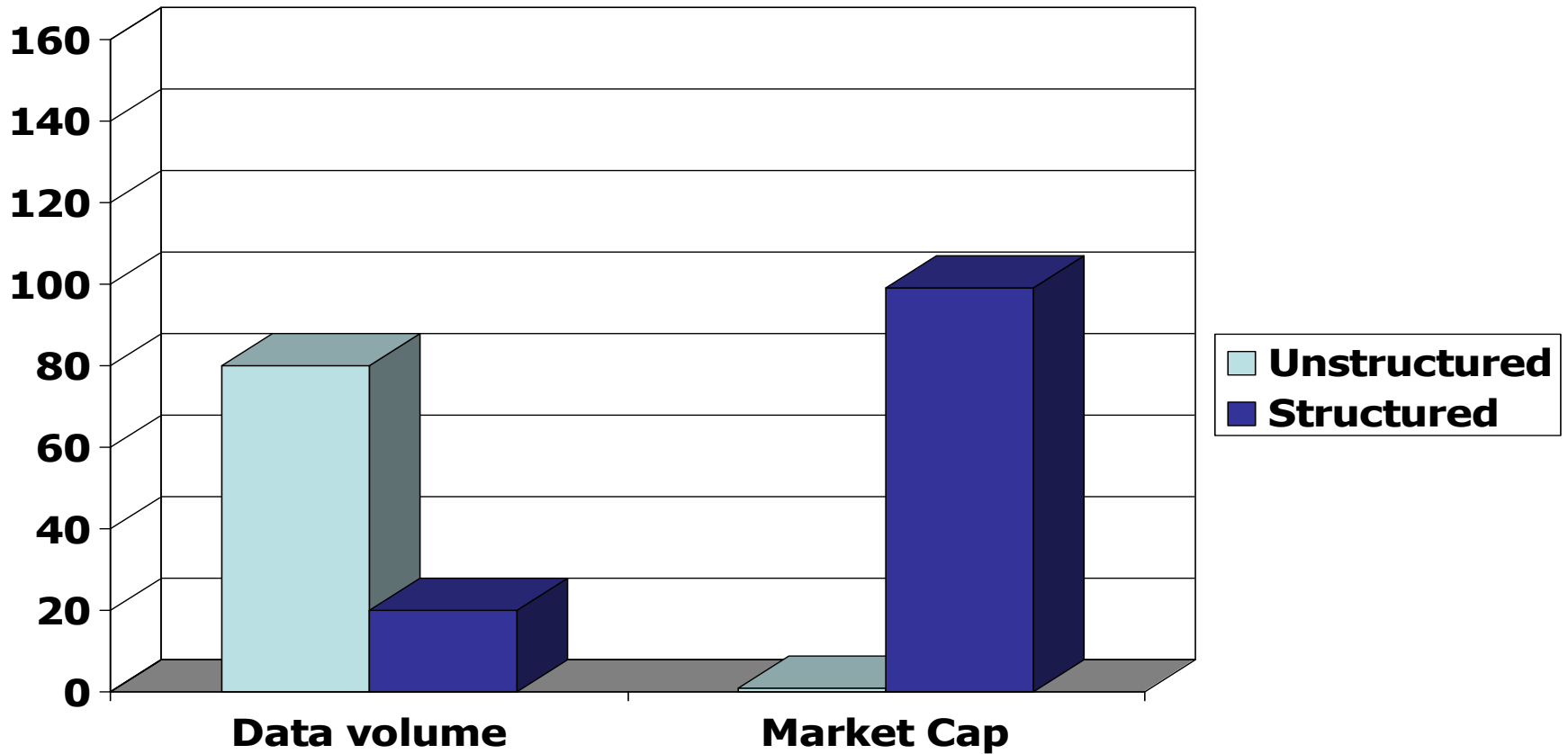
# I. Извличане на Информация (Information Retrieval)

- Намиране на документи, които удовлетворяват определена информационна нужда, в голяма колекция от документи.
- ~„Search Engine Technologies”
- Свързано е и с големи данни (Big data)

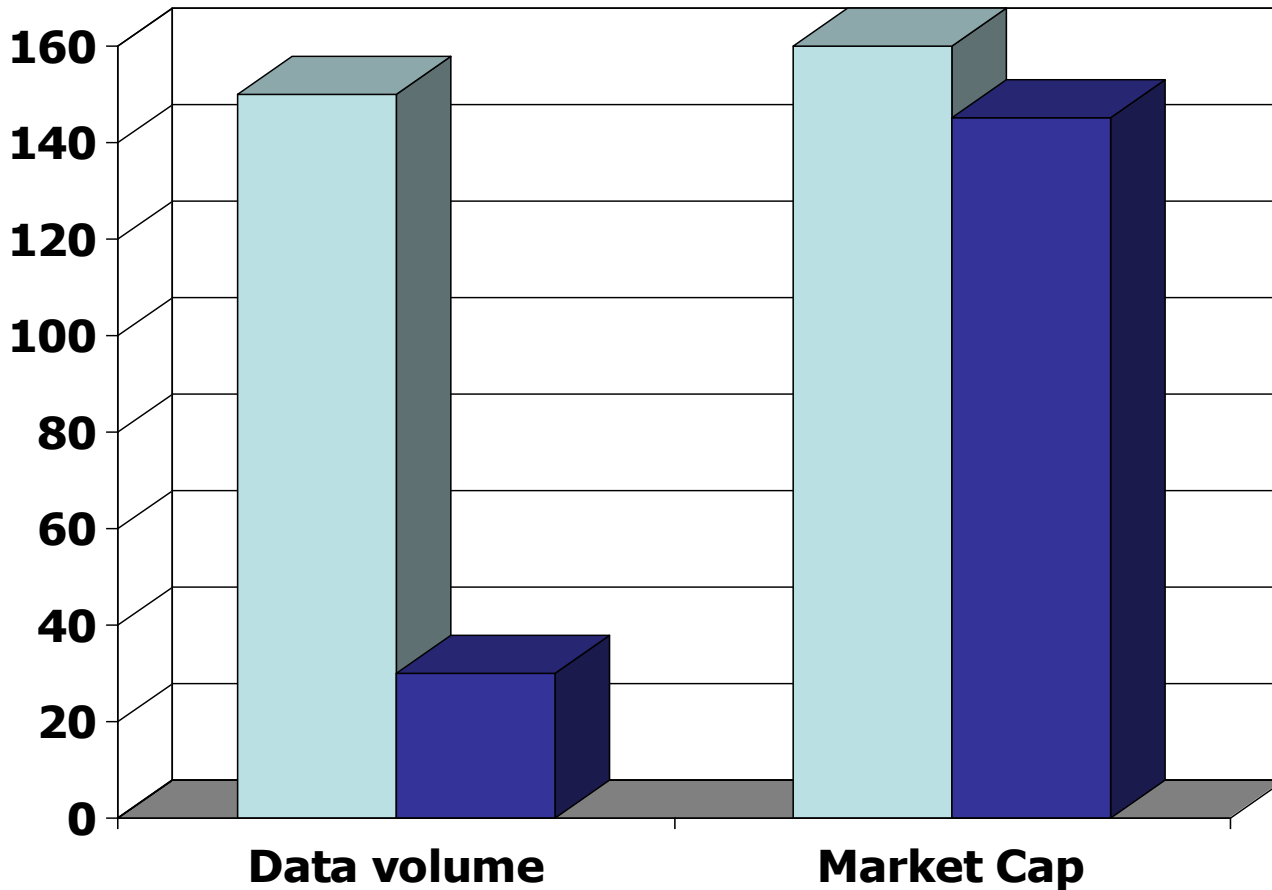
# Защо да уча “Извличане на информация” ?

- Какво ще науча?
  - Как се правят системите за търсене и намиране на документи в океана от неструктурирана информация (текст)
  - Компютърна обработка на естествен език
- 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



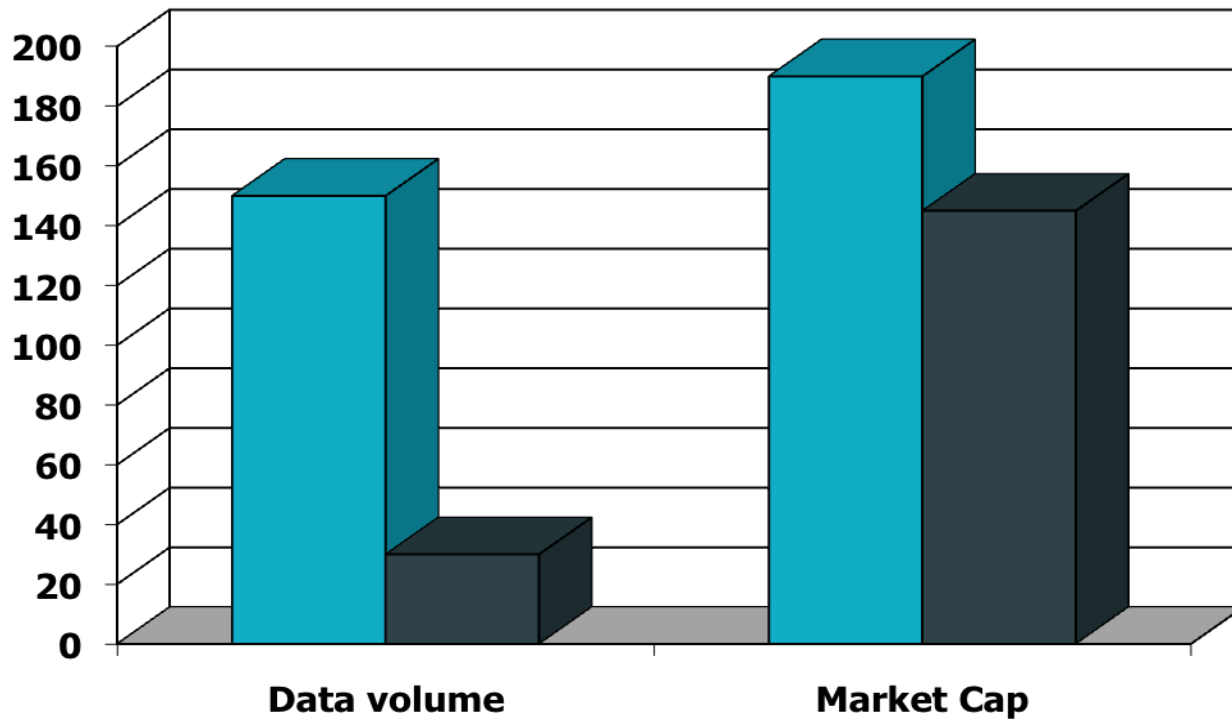
Google™

YAHOO!®

Unstructured  
Structured



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



Google™

YAHOO!®

■ Unstructured  
■ Structured

bing™

Ask™  
.com

# II. Откриване на знания (Knowledge Discovery (KD))

- Откриване на знания в данни  
(KD form Data, Data Mining, Data Science  
~~ BIG Data, Business Intelligence, Predictive ...)
- Откриване на знания в текст  
(KD form Text, Text Mining, Information Extraction, ...)
- Hot Topics: Web Mining, Opinion Mining,  
Sentiment Analysis, Automated fact checking...

# Защо да уча “откриване на знания”?

- Обществото произвежда огромно количество данни и текст (BIG Data!)
  - Източници: бизнес, наука, медицина, околната среда, социален уеб, спорт, ...
  - Потенциално ценен ресурс
  - Но суровите данни са непотребни!
- Нужни са техники за автоматично откриване на знанията от тях.

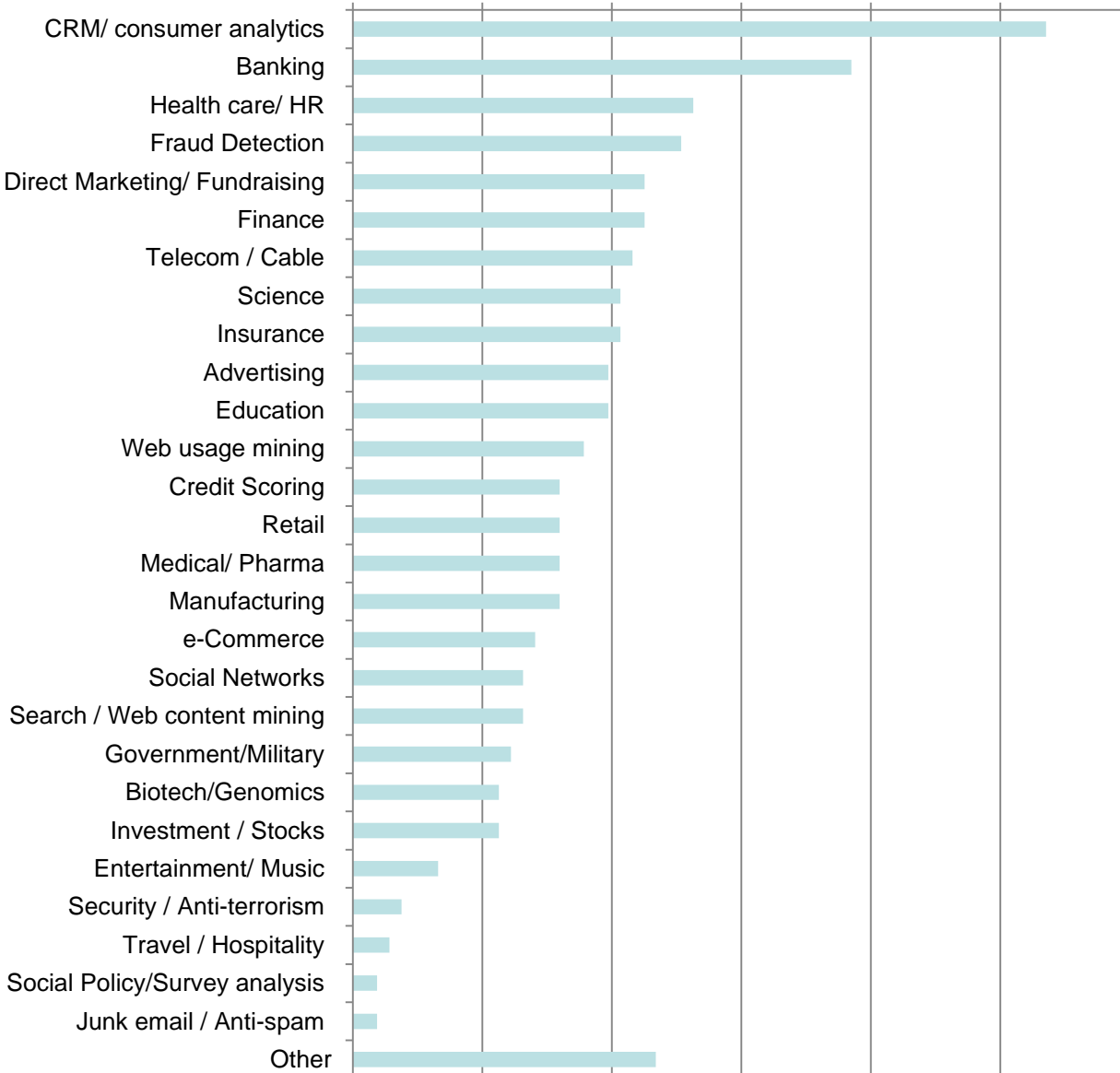


# Ще има ли пазар за мен?

- Data Scientist is the best job of the 21st century - Harvard Business Review
- Global Big Data market to reach \$122B in revenue by 2025 – Frost & Sullivan

# 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** (проф. И. Койчев и екип)
  - **Подходи за обработка на естествен език**  
(проф. Галя Ангелова и д-р Ивелина Николова)
  - ***Изкуствен интелект*** – ако не е вземат подобен като бак.
- **Летен семестър**
  - **Откриване на знания в данни**  
(проф. И. Койчев и екип)
  - **Откриване на знания в текст**  
(ст.н.с. д-р Преслав Наков и екип)

# Избираеми курсове

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

# Благодаря Ви за вниманието

За информация и въпроси:

проф. Иван Койчев (р-л на програмата)

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