NARRATOR
SYSTEM FOR REPORT GENERATION
IN NATURAL LANGUAGE

Inna Novalija, Marko Grobelnik
Artificial Intelligence Laboratory
Jožef Stefan Institute
Introduction

• **Goal** of the narrative reporting is to analyze the data and to present them in a simple and understandable way to the user.

• **Motivation** for automatic developing of stories in natural language – to minimize human efforts and funding, to obtain interesting conclusions from the data, to create compelling entertaining content in short period of time etc.

• **NARRATOR** is a system for report generation in natural language:
  o website performance analysis
  o data coming from Google Analytics services
  o report generation in English
  o general and user specific information
System Architecture

Google Analytics API
- mobile analysis
- content analysis
- advertising analysis

Indicators
- key performance features
- 126 predefined indicators
- possibility to combine
  - Examples:
    - number of visits last week
    - number of pageviews last week
    - most popular keyword last week
    - average time on site last week

Rules
- triggers
- contain indicators, numerical and logical operators
  - Examples:
    - [number of visits last week] > [number of visits previous week]

Templates
- combine together indicators, rules and statements in natural language
  - Examples:
    - <text>
      Traffic was down last week from the previous week
    </text>
    - <rule>
      [number of visits last week] < [number of visits previous week]
    </rule>

Report Scenarios
- contain the references to the specific template (sets)
  - Examples:
    - A general weekly report can provide information about traffic (traffic was up, traffic was down, traffic stayed the same) in the last week with comparison to the previous week
System Interface

Generated report for videolecures.net

YOUR GENERATED REPORT:
Traffic was up last week from the previous week.
Number of pageviews last week is up from previous week to 78467.0.
Engagement was up last week from the previous week.
Biggest source (by visits) last week was google with 17333.0 visits.
Data Mining Techniques

- **Decision tree** learning. Input - list of variables, output - model that predicts a value of the target variable.

- **Experiment**: building an experimental decision tree for the website videolectures.net.

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Attribute Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>isMobile</td>
<td>YES, NO</td>
</tr>
<tr>
<td>continent</td>
<td>AFRICA, AMERICAS, ASIA, EUROPE, OCEANIA, NOT_SET</td>
</tr>
<tr>
<td>visitorType</td>
<td>NEW_VISITOR, RETURNING_VISITOR</td>
</tr>
<tr>
<td>dayOfWeek</td>
<td>0 (Sunday), 1, 2, 3, 4, 5, 6</td>
</tr>
<tr>
<td>avgTimeOnSite</td>
<td>0_10, 10_100, 100_500, 500_AND_MORE (seconds)</td>
</tr>
</tbody>
</table>

**isMobile = YES**
- continent = AFRICA
  - dayOfWeek <= 3
    - dayOfWeek <= 0: 100_TO_500
    - dayOfWeek > 0: 0_TO_10
  - dayOfWeek > 3: 10_TO_100
- continent = AMERICAS: 10_TO_100
- continent = ASIA
  - dayOfWeek <= 4: 10_TO_100
  - dayOfWeek > 4: 100_TO_500
- continent = EUROPE: 100_TO_500
- continent = OCEANIA: 10_TO_100
- continent = NOT_SET
  - dayOfWeek <= 0: 100_TO_500
  - dayOfWeek > 0: 10_TO_100

**isMobile = NO**
- visitorType = NEW_VISITOR: 100_TO_500
- visitorType = RETURNING_VISITOR: 500_AND_MORE
Conclusion

- **NARRATOR** is a system providing the possibility to **transform the numerical data** from Google Analytics services into reports in natural language.

- **NARRATOR** provides a set of report scenarios, based on numerical and textual indicators.

- **NARRATOR** uses Google Analytics data for a specific periods of time.

- Inside **NARRATOR** system, we are developing a functionality, which allows to get more detailed analysis of the website data.

- For the future work we consider the further development of the **NARRATOR data mining functionalities**, as well as the technology adaptation for other data sources and data streams.
Demo

http://narrator.ijs.si
Questions?