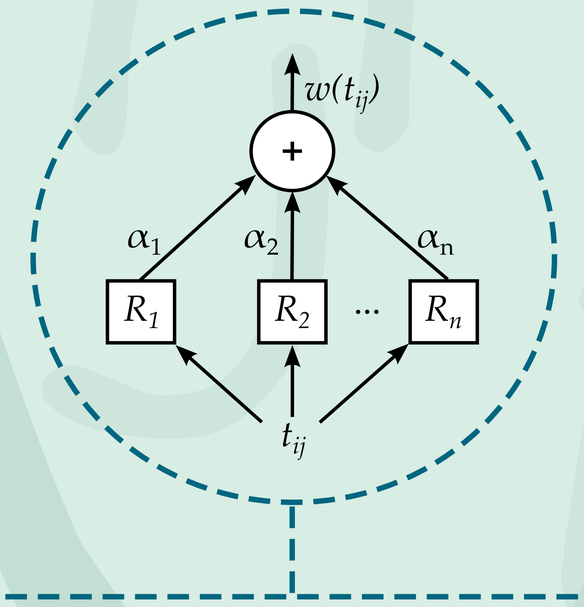
Personalized Text Summarization

Róbert Móro

Supervisor: prof. Mária Bieliková

Motivation

- information overload problem
- conventional summarization methods consider only text of a summarized document
- however, we can utilize other sources of information:
  - users' characteristics
  - document metadata (expert- or user-added)



Pre-processing

- machine translation to a reference language
- tokenization and terms extraction
- sentence segmentation

Construction of a personalized terms-sentences matrix

$$w(t_{ij}) = \sum_k \alpha_k R_k(t_{ij})$$

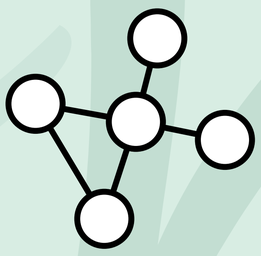
Terms frequency and terms location rater

Annotations rater

- user-added highlights indicating importance of text fragments
- extended by popular highlights

Relevant domain terms rater

- concepts identified by domain experts
- lightweight semantics captured in the domain model

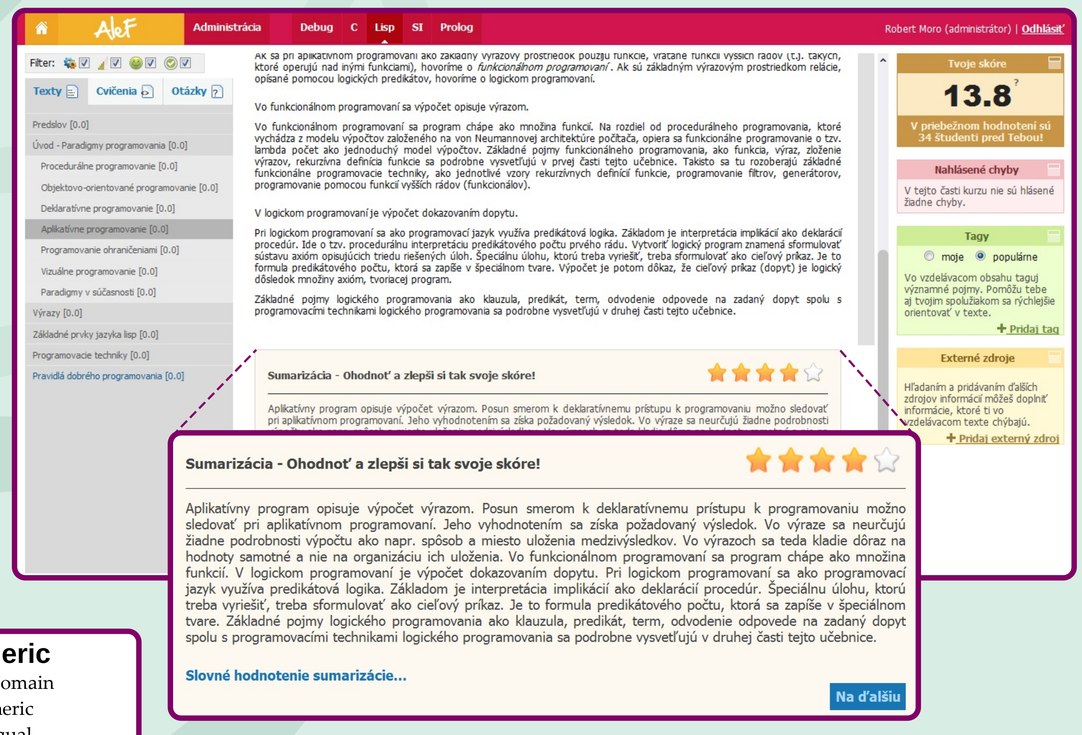


Knowledge rater

- actual user's knowledge of concepts
- uses user overlay model as a source

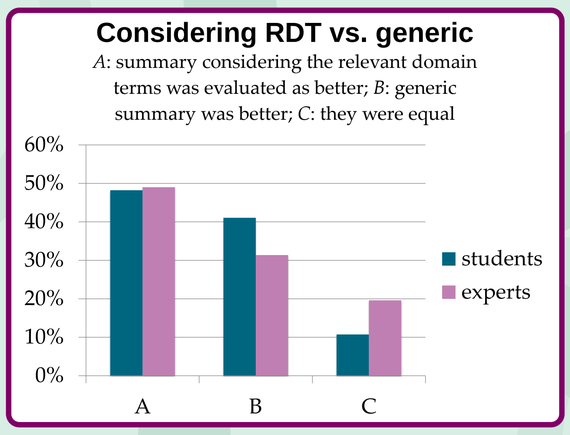
SVD and sentences selection

- sentences are selected from the original document (not translation)



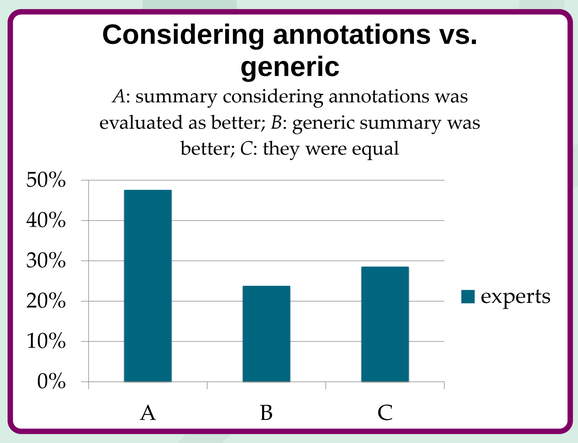
Experiments

- in the domain of learning (knowledge revision scenario)
- 2 experiments: FLP and PSI course in ALEF
  - 75 students, 5 experts, 303 LOs summarized
  - 2242 ratings, 385 comparisons, 479 answers to evaluation questions



Evaluation

- summarization considering relevant terms gained on average higher score than generic (*3.79* vs. *3.54*)
- expert comparison of variants gave us similar results, also when considering annotations
- summaries capable of summarizing important concepts, even for revision



Contributions

- our approach is domain- and language-indendent
- our contribution lies in the proposal of
  - the specific raters that take into account terms relevant for the domain or the level of knowledge of an individual user
  - the method of the raters' combination which allows considering various parameters or context of the summarization

STU FIIT SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLÓGIÍ

xmoror@stuba.sk

PeWe@FIIT personalized web group