

Extracting Keywords from Educational Content

Jozef Harinek

Supervisor: Marián Šimko

Motivation

- Many educational documents
- Attached annotations
- ...why not get use of it?

State of the art

- Domain model automated creation [1]
- Tweets utilized for keywords extraction from web resources [2]
- ...however we are not aware of works that study utilizing user-created annotations for keywords extraction

Goals

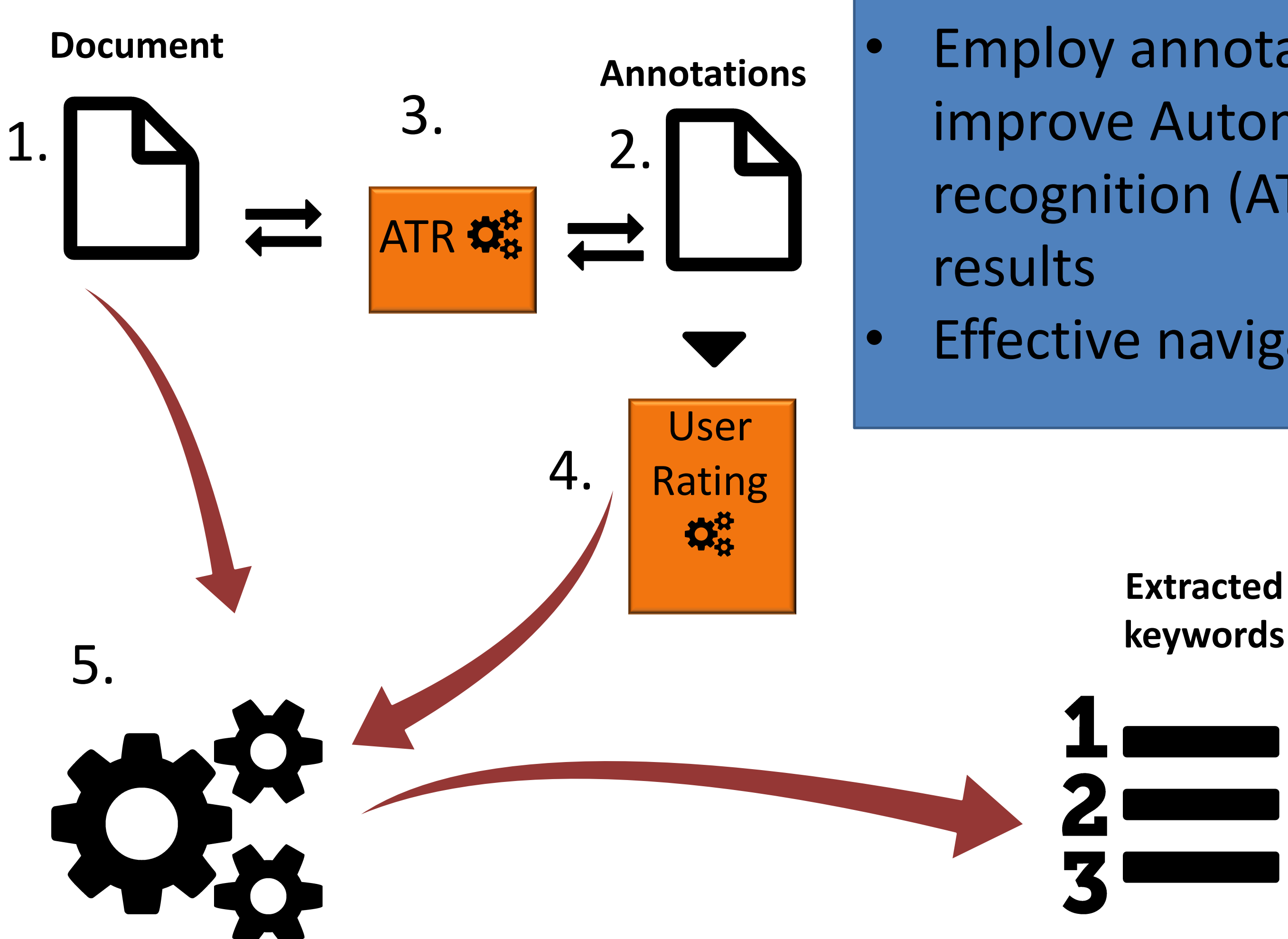
- Employ annotations and improve Automatic term recognition (ATR) algorithms results
- Effective navigation

Method description

1. Document text preprocessing
 - lemmatization , stop words removal
2. Annotations preprocessing
 - a) creating Extended document of annotations
 - b) lemmatization, stop words removal
3. Extract keywords with Automatic term recognition algorithm
4. Include User Rating in annotations
5. Combine two sets of keywords

$$w_{annot} = \sum_{a \in A} \sigma_a \sum_{u \in U_t} UserRating(u) * w_{ATR}(t, E_d)$$

$$w_{final} = (1 - p) * w_{ATR} + p * w_{annot}(t, d)$$



Conclusions

- Mainly tags, highlights, and basic comments processing
- Keywords extraction with user-added annotations processing is reasonable (52 % improvement)
- Good for short educational documents with user-added annotations
- Still some work to be done:
 - More advanced comments processing
 - Various ATR algorithms
 - Various User Ranking methods
 - Integration
 - Come2t
 - ALEF

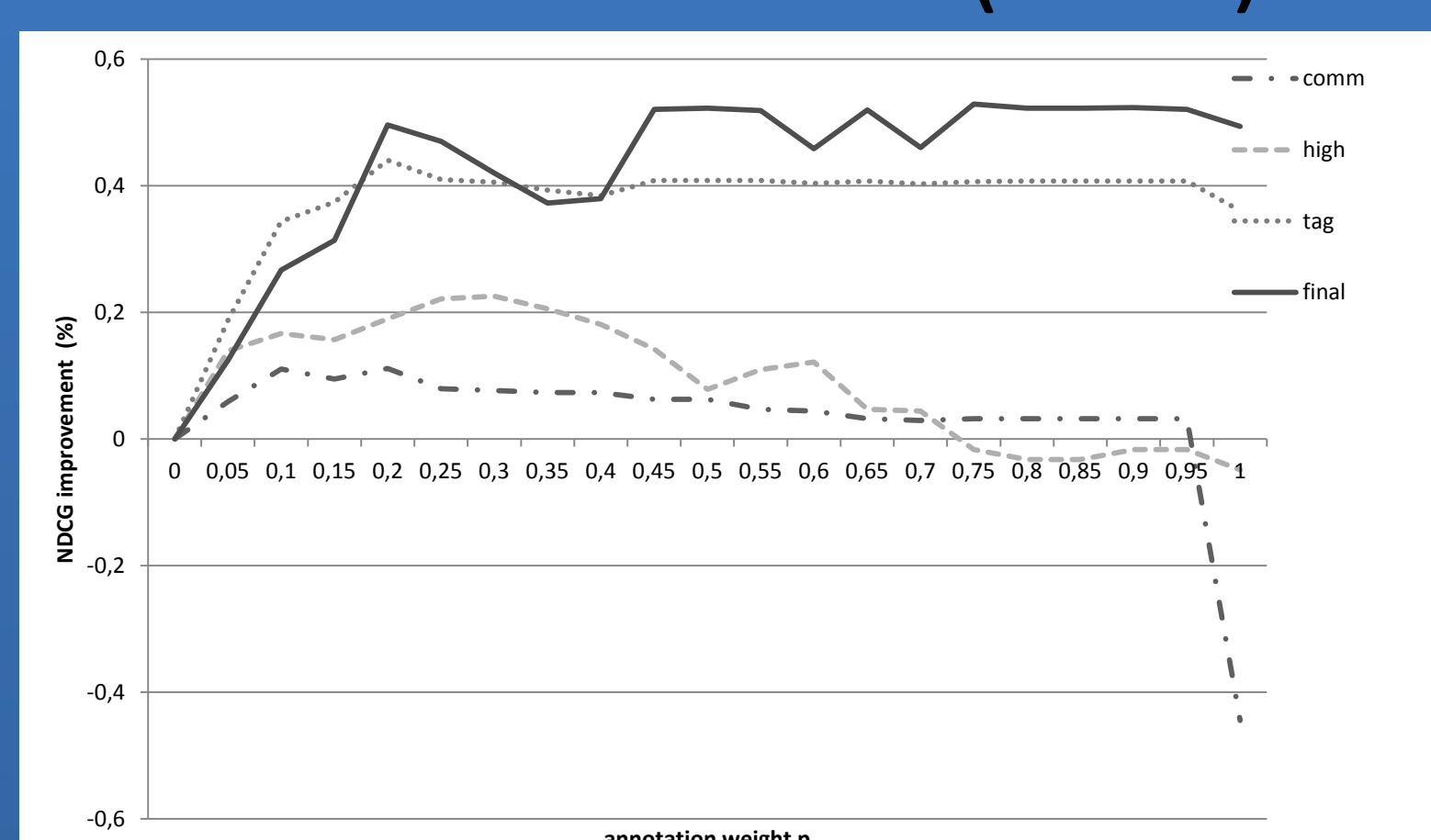
Experiment

Set up

- 185 LO – ALEF (PSI)
- 329 students
- 21,491 annotations
 - 5,348 tags
 - 15,707 highlights
 - 436 comments
- Golden standard
- Tags, Highlights, Comments

Results

- Annotations processing yields 52 % improvement in keywords extraction
- Most influential – Tags (44 %)
 - Highlights (22.5 %)
 - Comments (11 %)



References

[1] Šimko, M., Bieliková, M. 2009. Automated Educational Course Metadata Generation Based on Semantics Discovery. In LNCS 5794, *Proc. of European Conf. on Technology Enhanced Learning*, ECTEL 2009, Springer, pp. 99–105.

[2] Uherčík, T., Šimko, M., Bieliková, M. 2013. Utilizing Microblogs for Web Page Relevant Term Acquisition. In *SOFSEM 2013: Theory and Practice of Computer Science*. LNCS, Vol.7741, Springer, pp. 457–468.