

Lexicographic Data meet Computational Linguistics and Knowledge Systems

The ENeL WG3 meeting in Brno
September, 16-17 2016

But wait - What does this long Title mean?

Take, for example, machine translation:

- How does lexicography boost MT?
- How does MT improve lexicographic work?

Take lexicography and (lexical) linked (open) data, the semantic web, machine learning, knowledge systems, artificial intelligence

- Can this data be used to improve automatic conversion, e. g. going from XML to RDF
- Does the availability of data improve ML techniques

Is this all?

No. There is more:

- Content modelling and data structures that enhance interoperability with language technology
- Multi- and cross-lingual search, information retrieval, and data mining
- Web as corpus and lexicography
- Transformation technologies, like text2speech, OCR

But Wait - XYZ was missing...

Well, a potpourri of our topic ingredients looks like this:

machine translation (MT), e-learning, word processing, spell checker, grammar checker, search engines, query expansion, search engine optimisation (SEO), text2speech, speech2text, optical character recognition (OCR), big data, knowledge systems, linked data, semantic web, internet of things, ...

But Wait - XYZ was missing...

..., artificial intelligence, machine learning (ML), sentiment analysis, data mining, web as corpus (WaC), question answering, automatic summarization, word-sense disambiguation (WSD), thesauri creation and expansion, part-of-speech (POS) tagging, multilingualism, domain adaptation, named entity recognition (NER), coreference resolution, semantic parsing, ontologies, terminologies

And you can add more!

Questions

Starting from technologies/problems/solutions, in any of the relevant domains, we could think of

- What does this bring to the table of a cross-domain undertaking?
- How can we leverage 'knowledge' from the other domain?
- What and how can lexicography contribute to other domains?

Different Perspectives onto the Potpourri



Different Perspectives onto the Potpourri

- ★ Scientific Fields (Sciences)
 - MT, e-Learning, Knowledge Systems, ML

- ★ Service-like technologies (not applications but services)
 - MT (as a service), editorial work (spell, grammar, etc. checking), SEO, OCR, data analytics (data mining et al.), terminological database

Different Perspectives onto the Potpourri

- ★ Technologies (Research questions with industry level applications)
 - MT, e-learning (platforms), word processing applications (including spell checker, grammar checker), search engines, text2speech, speech2text, OCR, ML, question answering systems, automatic summarization tools, NLP pipelines (tokenisation, lemmatisation, POS tagging, WSD, NER, coreference resolution, semantic parsing)