

Zápis
z veřejné odborné rozpravy k výsledkům projektu Centrum počítačnické
lingvistiky
konané 30. listopadu 2004 v Praze

Protocol
on public scientific discussion about outcomes of the project Center for
Computational Linguistics
held in Prague, November 30, 2004

Přítomni: dle prezenční listiny

Členové hodnotícího panelu (bez titulů, vesměs zahraniční profesoři): Eduard Hovy (předseda), Nicoletta Calzolari, Bonnie Dorr, Frederick Jelinek, Aravind Joshi, Owen Rambow, Louise Guthrie

Program:

A. Presentations of the members of the research team

1. Presentation of the Prague Dependency Treebank (Eva Hajičová, Jan Hajič):
 - a. Overall characterization
 - b. The annotation of the surface shape of the sentence: morphology, analytic layer
 - c. Tectogrammatical annotation: underlying syntactic structure, topic-focus articulation, coreference
 - d. Perspectives
2. Software tools developed during the project and available for broader usage (Jan Hajič)
3. Valency dictionary (Markéta Lopatková)
4. Search in PDT (Veronika Kolářová)

B. Panel discussion: Panel chair: Eduard Hovy

Members of the panel: Nicoletta Calzolari, Bonnie Dorr, Frederick Jelinek, Aravind Joshi, Owen Rambow, Louise Guthrie

Conclusions of the panel discussion:

The panel considered the achievements and impacts of the project. Each panelist was given the following four questions:

- What are the major accomplishments and contributions?
- Where is resource-oriented NLP headed in general?
- Please try to measure the success of the project
- What are the future directions (next 5 years)?

Their discussion rated the project along several different dimensions, which can be organized into two categories: short-term impact (publications, resources, and education), and longer-

term impact (the legacy: work by others building upon it, and change in directions of general (international) research). All panelists agreed that the project has made numerous significant achievements, most of them at the level of the best comparable international work, and in several cases going beyond it. For example, the cornerstone product, the Prague Treebank, contains more information and is more carefully constructed than the Penn Treebank (which is the most-used treebank in the world today). The associated tools (parsers, editors, taggers, linguists' search engines, etc.), and the work on speech recognition, is thoroughly executed and has in some cases (like the tree editor) already been used by some panelists in their own work. The project's numerous publications appear in the top international conferences and journals, and are comprehensively listed on the website.

The panel unanimously felt that it is a pity that this work must now stop, and recommended several steps to avoid the results being lost:

- Project members should organize tutorials and symposia in the near future in order to share their accumulated experience to educate people in other countries
- All additional resources should be made available to international resource institutions such as LDC, for continued distribution
- Project members should initiate and/or participate in international consortia for ongoing annotation of texts in other languages, again to preserve their experience
- If possible the Prague Treebank should be translated/converted into other languages and distributed

In Prague, November 30, 2004

For the panel:

Eduard Hovy