Andrej Gardoň
Natural Language Processing Centre

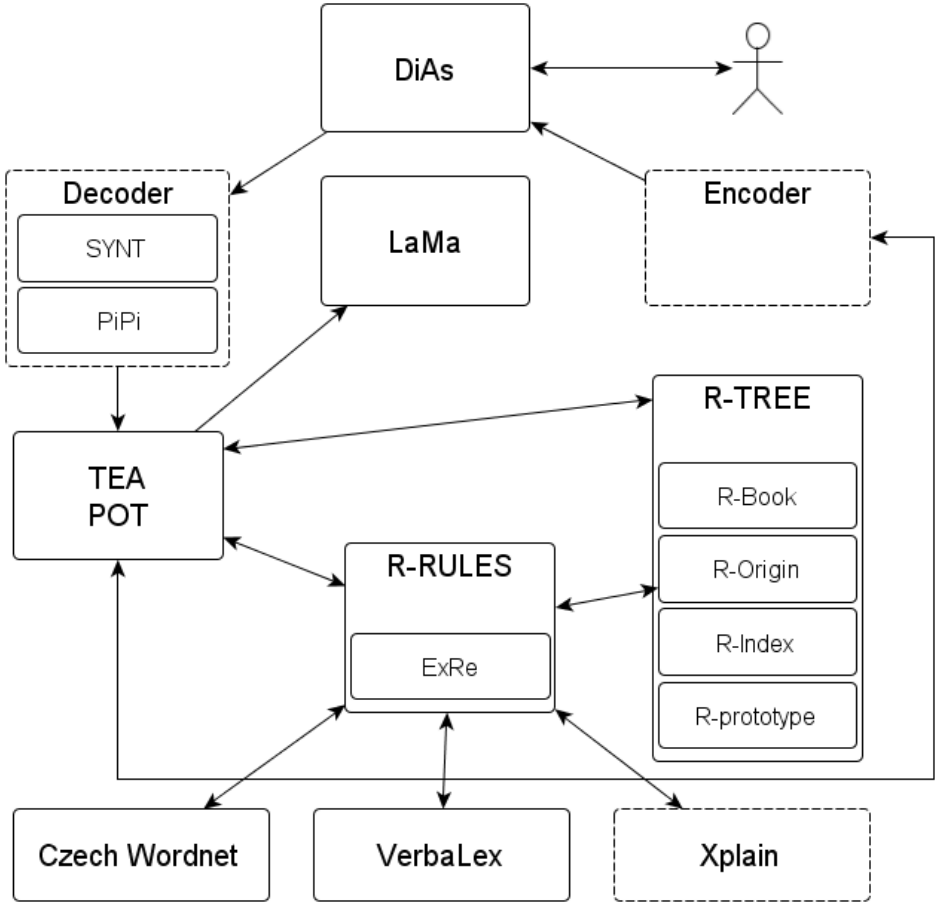
CZECH KNOWLEDGE-BASED SYSTEM WITH TEMPORAL REASONING

Outline

- QA systems are usually oriented on English language
- Present QA systems have insufficient time capabilities
- Poor processing of repetitive actions
- No complex model for subsequent events and relations between them

Czech QA system based on TIL

- Long term project, interconnection of several projects



TIL KB

- Module based:
  - DiAs – manages dialogue with a user
  - Decoder – utilizes SYNT and Saara to transcribe Czech NL sentence into TIL formula
  - LaMa – language manager, keeps mapping between words of a language and internal representation of a referent (in TIL context)
  - TEA POT – translates TIL formula into language independent R-TIL representation (using LAMA). Communicates with other modules – processing unit
  - R-TIL – ex. O(Apple) in TIL eq. O(@256) in R-TIL
  - R-TREE – KB storing R-TIL formulas. Supports prototype generation and keeps the possible R-TIL formulas constructing given referent
  - R-RULES – inference engine with API for external resources (VerbaLex, WordNet…)
  - Encoder – from R-TIL to TIL and Czech NL sentence

Thesis Goals

- Implementation of basic components (LAMA, TEA POT, R-TREE, R-RULES)
- Incorporating of solutions to problems identified during previous research on Dolphin
- Set of inference rules for time processing namely:
  - Direct time points and intervals (Andrej was born on 30th of November 1984);
  - Relative time points and intervals (Maria visited Andrej yesterday);
  - Undefined time points and intervals (Maria visited Andrej);
  - Repetitive activities (Andrej goes swimming every Thursday);
  - Interpretation of grammatical tenses
  - Relations between subsequent events and episodes.

Evaluation

- TIL corpus of sentences focused on presentation of Dolphin capabilities
- Part of the corpus based on news servers like tyden.cz, aktualne.cz
- Corpus parsed by Synt and translated into intermediate TIL corpus (contains alternative TIL formulas to a sentence)
- Manual selection and post processing of best TIL formulas provided by SYNT (final TIL corpus)
- User can choose model question to the Dolphin from web-like interface
- Direct input of TIL formula into the Dolphin system
- Output of Dolphin are TIL formulas