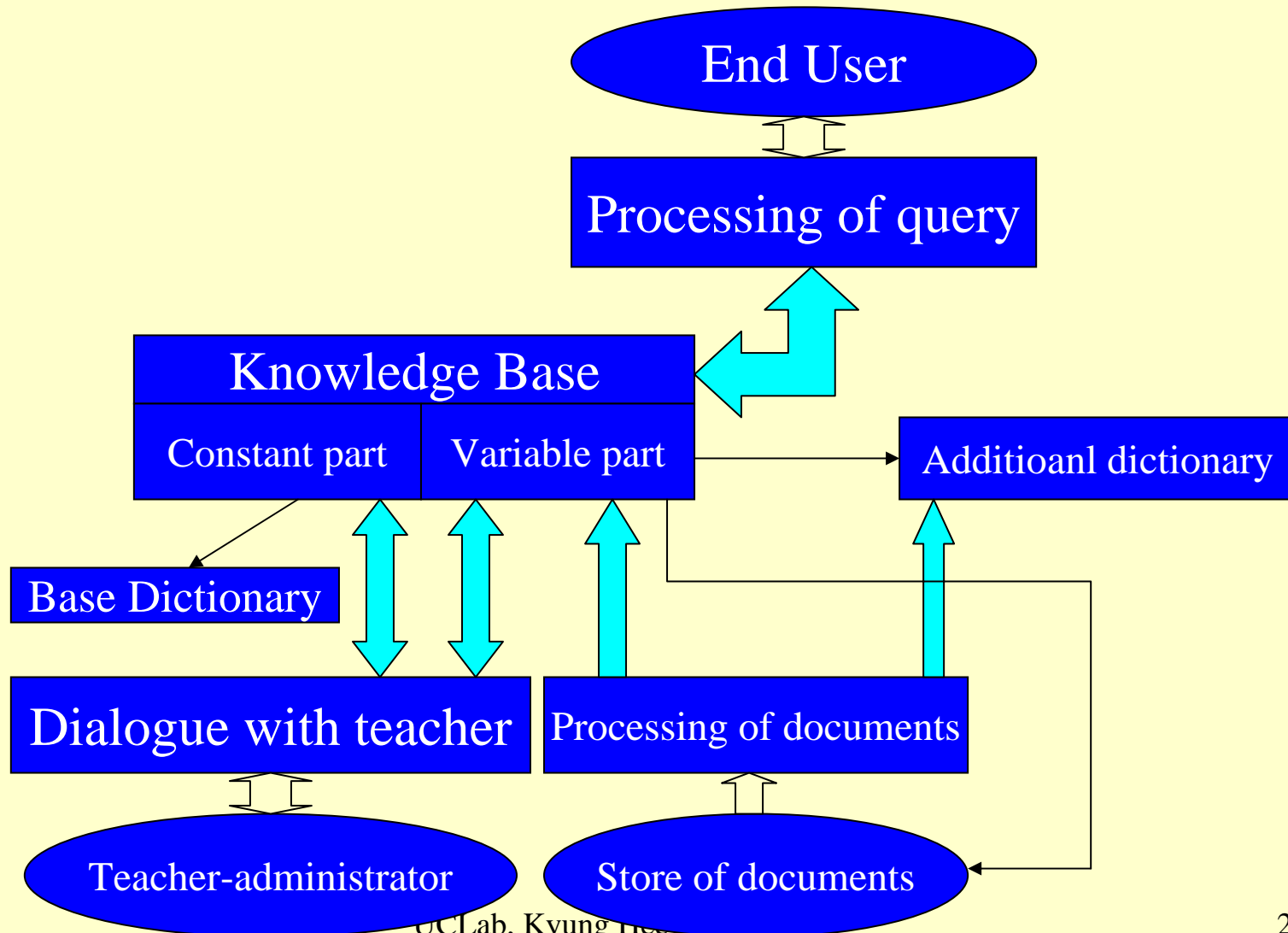


Hybrid Intelligent Systems

Lecture 14. Part 2

Example of hybrid approach in NLP for searching system

Technology for building of learned system for search of documents by sense



Main principles in proposed technology

- Orientation to recognition of semantics with minimum usage of knowledge about syntax of the language,
- Creation of hierarchies from concepts with horizontal (associative) links between nodes of these hierarchies as result of processing of documents,
- Recognition of words and word collocations on maximum resembling with usage of neural algorithms.

Kinds of frames

1. frame coupled immediately to the word or the document (the frame-word or the frame-document)
2. frame, with which associates a word collocation (composite frame)
3. frame-concept including the links on several other frames, playing the defined role in this concept

Slots of frame

- Parent - link on the frame - parent or class (vertical links)
- Owner - list of links to frames - concepts or the composite frame, in which structure enters the given frame
- Obj - object participating in concept,
- Subject - subject (or main object), participating in concept
- Act - operation (action) participating in concept
- Prop - property participating in concept
- Equal - list of concepts - synonyms circumscribed in the given frame (horizontal links)
- Unequal - list of concepts - antonyms circumscribed in the given frame
- Include - list of links to the frames switched on in the given concept constituent (vertical links)

Other main parameters of frame

- Level - level of the frame in hierarchy
- DocName - index of filename (path) of document coupled to the frame
- IndWord - index of a word in the dictionary coupled to the frame
- H - threshold of operation of the frame, as neuron
- Role - role of the frame in concept, which it enters or can enter (A-operation, O-object, S-subject, P-property, U-undefined or D - the operation at the analysis (by special procedure))
- NO - indication of inversion of the frame

Dictionaries

- Basic, in which the words with their roles (essence, operation or property, in other words - noun, verb or adjective are stored
- The supplemented (dynamic) dictionary including a words, not recognized in the base dictionary
- Dictionary of special words, associated with separators and analyzed as separators.

Steps of analyzing of sentence in context of learning

- 1) selection of words (using signs of punctuation and spaces)
- 2) the recognition of words on maximum resembling with words in the dictionary, thus if the approaching word is not in the fundamental dictionary, then searching of this word in the supplemented dictionary, and in fail case this word adds in this dictionary
- 3) the creation of the frames of a level 0, the result of this stage is object-sentence representing list of the frames
- 4) replacement in this object of special words by signs-separators,
- 5) processing of the object-sentence by a procedure of recognition-creation of the frames of levels 1 and 2

Steps of analyzing of sentence in context of processing of query

- 1) selection of words (using signs of punctuation and spaces),
- 2) the recognition of words on maximum resembling with words in the dictionaries. In case of unknown word system ask question "what is < new word >?". The answer of the user is processed in context of learning.
- 3) the creation of the frames of a level 0, the result of this stage is object-sentence representing list of the frames,
- 4) the recognition of the frames of a level 1 or 2 - word collocations in the knowledge base maximum similar to recognized phrase (here is used neural algorithm, i.e. weighed addition of signals from words, entering into the frame, or frames and matching with a threshold),
- 5) the searching associatively coupled by the links Equal with the recognized phrases of the frames (level 0), coupled with documents,
- 6) the searching of frames-documents from the retrieved frames on connections such as include, act, obj, subject, prop from above downwards
- 7) the output of the retrieved names of documents or words which are included in structure of the retrieved frames.

Steps of learning of System

- Initial tutoring to recognition of structure of sentence by input of sentences as "word - @symbol". This step provides creation of dictionary of special words
- Initial tutoring. During this step the knowledge base is filling by fundamental concepts from everyday practice or data domain as sentences such as "money - means of payment", "morals - rule of behavior", "kinds of business: trade, production, service" etc.
- Base tutoring. In this step the explanatory dictionary of data domain is processed, where the concepts of any area are explained with use "-" or corresponding words.
- Information filling. In this step the real documents are processed.

