

## **Individual Paper Proposal**

### **Abstract**

Introduction part of the paper is going to provide some reflections and justified information about reading and textual information concepts. Two concurrent contexts are reviewed: human brain and computer processor. This theoretical part is to help to understand the nature of information encoding, particularly in textual form.

Then the main goal of the base work will be defined: to encode human readable dictionary from the lexicographic point of view. And the second general goal: to achieve XML/TEI P5 standard which is proposed as a way towards the global data interchange (TextGrid).

Methodology part introduce dictionary as a hierarchical data structure presented in human perceptible/readable form. Where rules for the structure interpretation rather understood intuitively or, if required, available explicitly in the description part of the dictionary. Based on this rules the structural schema (lexicographic schema) is created which is the road-map for text's logical structure encoding. Encoding procedure involves structural elements recognition and their markup, which are to be implemented programmatically. Implementation part describes top-down approach, when elements in the top of the hierarchy (larger ones) processed in the first order and produce output which is passed to the next level processor which in turn performs recognition and tokenization of the next level elements (smaller ones). This approach keeps logic clear and thus reduce the complexity.

Technological aspect is based on CSS parser in conjunction with XML/HTML parser. Any textual element accessed via markup language parser and its style via mapped style parser. This technological background provides all necessary control and manipulation opportunities for textual information processing.

Output of the parsing is a machine-readable XML/HTML tree which reflects dictionary logic and can be processed by means of standard markup language processors, e.g. XSLT. Statistical results of the parsed dictionary will be provided in the conclusion section.