Paradigmatic and Syntagmatic Relations in Information Systems over Ontological Graphs

Krzysztof Pancerz[1,2]

[1] University of Management and Administration
Akademicka Str. 4, 22-400 Zamość, Poland
kpancerz@wszia.edu.pl
[2] University of Information Technology and Management
Sucharskiego Str. 2, 35-225 Rzeszów, Poland

People make sense of a text by identifying semantic relations which connect the entities or concepts described by a text (cf. [2]). Therefore, in the search for smarter, more human-like, computer tools, we need to equip such tools with ability to identify and utilize semantic relations in processing the texts. In [4], we have dealt with the problem of mining real-estate listings. In this problem, a special attention should be focused on preprocessing steps transforming advertisements in the textual form into information systems. Information systems have been proposed by Z. Pawlak as knowledge representation systems (cf. [5]). Any information system can be represented in a tabular form, i.e., as a data table called information table. Columns of the table are labeled with attributes, rows are labeled with objects, and entries of the table are values of the information function. Many machine learning and computational intelligence methods and algorithms, especially those based on rough set theory, need to capture and represent data in a machine-friendly format. Therefore, they work with data stored in information tables.

In the paper, we would like to show the main idea of incorporating paradigmatic and syntagmatic relations into mining data stored in information tables. In [6], F. de Saussure distinguished between syntagmatic and paradigmatic (associative) relations. Paradigmatic relations hold between concepts belonging to the same grammatical category. Paradigmatic relations cover a wide variety of associations between words, including morphological and phonetic. Syntagmatic relations hold between two or more words co-present in a sequence. Combinations based on sequentiality are called syntagmas. The notion of a syntagma applies among others to group of words and to complex units of every size and kind, for example, phrases, sentences.

In [3], ontologies, represented by means of graph structures (ontological graphs), were incorporated into information systems to deal with paradigmatic relations in data mining processes. In the ontological graph, each node represents one concept from a given ontology, whereas each edge represents a semantic relation between two concepts. There are a lot of paradigmatic relations defined in the literature. For example, WordNet [1] represents around a dozen paradigmatic relations between concepts, including: synonymy, antonymy, hyponymy, hyperonymy, meronymy, and holonymy. Information systems, in which attribute values are concepts from ontological graphs assigned to attributes, are called

simple information systems over ontological graphs (see [3]). We can use simple information systems over ontological graphs to describe syntagma data. Each vector of consecutive concepts in the row of the information table, representing the information system, represents one syntagma. Therefore, a simple information system is called a syntagma information system. Each object in such a system is said to be a syntagma. An n-tuple of attributes is called a scheme of syntagmas. It is worth noting that, in case of information tables, paradigmatic relations should be considered vertically (according to attributes), whereas syntagmatic relations should be considered horizontally (according to objects). One can see that, in syntagmas, succession of concepts is important. Each concept in a given syntagma, excluding the last one, can anticipate the next concept. Some ambiguities in anticipation of concepts in syntagma information systems can be described using the rough set approach. Analogously to rough set theory [5], we define the lower and upper syntagmatic anticipation of a given concept.

References

1. Fellbaum, C. (ed.): WordNet - An Electronic Lexical Database. MIT Press (1998)
2. Nastase, V., Nakov, P., Séaghdha, D.O., Szpakowicz, S.: Semantic Relations Between Nominals. Morgan & Claypool Publishers (2013)
3. Pancerz, K.: Toward information systems over ontological graphs. In: Yao, J., et al. (eds.) Rough Sets and Current Trends in Computing, LNAI, vol. 7413, pp. 243–248. Springer-Verlag, Berlin Heidelberg (2012)
4. Pancerz, K., Mich, O.: Mining real-estate listings based on decision systems over ontological graphs: Extended abstract. In: Popova-Zeugmann, L. (ed.) Proc. of the CS&P'2014. pp. 176–179. Chemnitz, Germany (2014)
5. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
6. de Saussure, F.: Course in General Linguistics. Open Court, Chicago and La Salle (1986)