**Generating and Using a Qualified Medical Knowledge Graph for Patient Cohort Retrieval from Big Clinical Electroencephalography** (**EEG) Data**

Sanda M. Harabagiu, PhD1

1Human Language Technology Research Institute and Department of Computer Science, University of Texas at Dallas, U.S.A.

Natural language processing of the narratives from EEG reports enables the recognition of EEG events and EEG activities as well as their clinical correlations, the patient state and the patient’s clinical picture. To organize these medical concepts we have designed an EEG-focused Qualified Medical Knowledge Graph (EEG-QMKG) in which concepts (e.g. EEG events, EEG activities) are qualified by their *modality* (e.g. factual, conditional) and *polarity* (e,g, positive or negative). The qualification of concepts takes into account the hedging used by physicians in EEG reports. Also, this representation events are grounded *spatially* (e.g. “frontocentral”) and *temporally* (“before” seizure) and all concepts are associated with their attributes (e.g. “small” amount of theta). Since many of the events and artifacts as well as medical concepts are mentioned multiple times in the same EEG report, co-reference was resolved automatically in order to best capture the context of medical concepts. The EEG-QMKG also includes the results of automatically processing the EEG signals interpreted in reports.

By capturing automatically the semantics and section-informed cohesion from the EEG reports, we designed the EEG-QMKG as a probabilistic representation, in which edges between nodes are inferred statistically through BigData techniques, such as MapReduce. Active learning controls the quality of the knowledge captured in the EEG-QMKG. The EEG-QMKG is a central component our architecture for retrieving patient cohorts when queries such as “*young patients with focal cerebral dysfunction treated with Topamax*“are being posed. Not only does the EEG-QMKG enable the retrieval of a list of ranked EEG reports and signals that satisfy the inclusion criteria set by the query, but it also informs innovative learning-to-rank techniques based on results of acceptance testing using three focus groups: expert annotators, clinicians and medical students. The learning-to-rank framework will produce optimal patient cohorts and contribute to regularization of interpretations in EEG reports.

Submitting Author’s Career Stage: Professor