

**Ph.D. Dissertation Defense**

**Department of Computer Science**

**& Engineering**

**“A STOCHASTIC PETRI NET BASED NLU SCHEME FOR TECHNICAL DOCUMENTS UNDERSTANDING”**

***By Adamantia Psarologou***

**Friday, April 29, 2016 – 12:00 pm**

**304 Russ**

**Ph.D. Committee:** **Drs. Nikolaos Bourbakis, Advisor, Soon Chung, Yong Pei, and Sukarno Mertoguno (Office of Naval Research)**

**ABSTRACT:**

Natural Language Understanding (NLU) is a very old research field, which deals with machine reading comprehension. Despite the many years of work and the numerous accomplishments by several researchers in the field, there is still place for significant improvements.

Here, our goal is to develop a novel NLU methodology for detecting and extracting event/action associations in technical documents. In order to achieve this goal we present a synergy of methods (Kernel extraction, Formal Language Modeling, Stochastic Petri-nets (SPN) mapping and Event Representation via SPN graph synthesis). In particular, the basic meaning of a natural language sentence is given by its kernel (*Agent → Action → Patient*), which is *“who”, is doing “what”, to “whom”*. Thus, we have developed a methodology that automatically extracts the kernels of NL sentences based on their parse trees. Then, we represent the kernel’s structure in a form of a formal language, called *Glossa*, for efficient processing. Next, we map the formal representation of kernels to an SPN state machine in order to embed timing in the representation of NL sentences. Finally, we synthesize the SPN representation of kernels for expressing the association of events/actions of different sentences.

Results of our methodology are presented to prove the concept and validate the overall approach. Moreover, we provide two different application that our proposed NLU methodology can be used for. The first application is a quick and easy way for modifying technical documents, by multiple users. The second one is document summarization, where two different types of summarization are described.