**Adaptive Information Extraction (Turmo, et al, 2006)**

Turmo and his team of researchers have presented an adaptive Information Extraction system that can be used to extract information from different type of texts like unstructured, structured and semi-structured texts. In their paper, they presented the architecture that they used in building the system. Turmo’s IE system has six main modules in its architecture. Module 1 is responsible for the named-entity recognition part of the IE system. For this module, Turmo et al used a commercial tool called NetOwl Extractor 3.0 to recognize general named-entity types. It is in this module where time/numerical expressions, names (persons/places/organizations), acronyms (organization names/locations), and semantic subtypes (country/city) are being recognized and extracted. Moving on, Module 2 or the Custom NameTag module is responsible for the recognition of restricted-domain named-entities by using pattern matching. The output phrases for this module are SGML-tagged (Standardized Generalized Markup Language) into the same input document. On the other hand, Modules 3 & 4 are responsible for SGML-tagging the phrases in the sentences that are considered to be values for the slots defined in the templates and works hand-in-hand. Module 3 or the PhraseTag module works by applying syntactico-semantic rules identify the noun phrases in the previously recognized/extracted named-entities. Module 4 or the EventTag module works by applying a set of custom-built syntactico-semantic multi-slot rules to recognize/extract events from the input sentence. Moving on, Module 5 or the Discourse Analysis Module is responsible for coreference resolution or the merging of the previously extracted noun phrases. This module was implemented by using three different strategies so that it can be modified to reach optimal performance regardless of the extraction scenario. Strategy A or the Rule-Based Strategy uses a set of custom-built rules to resolve definite noun phrases and singular personal pronoun coreference. Strategy B or the Machine Learning-Based Strategy uses a decision tree that has been formed from learning a corpus tagged with coreferences. Strategy C or the Hybrid Strategy uses Strategy A to filter false antecedents and then uses Strategy B to rank the remaining antecedents. In general, Module 5 is just merging the partial templates formed by the previous module. Lastly, Module 6 or the TempGen Module is responsible for the completion of the templates generated from the previous module by considering the consistency of the values in the slots of the event templates after resolving the noun phrases conferences and the generation of the output in the desired format. Figure 3.1.1d illustrates the architecture of the system proposed by Turmo et al.

Figure 3.1.1d: The Architecture of Turmo et al.’s Adaptive Information Extraction System