**Name of the paper:** From user requirements to UML class diagram.

\*Hatem Herchi,

\*\*Wahiba Ben Abdessalem

\*Department of computer science, High Institute of Management of Tunis, Tunisia

\*\*Department of computer science, High Institute of Management of Tunis, Tunisia

**Existing tools/methodologies mentioned:**

* Ambriola and Gervasi proposed a framework to automatically transform user’s requirements into different models.
* The models included Data Flow graphs (DFG), entity-relationship diagrams, UML diagrams.
* The system takes as an input the problem’s description written in natural language.
* Then a domain based parser, called CICO, and is used to extract some facts from them.
* Zhou and Zhou implemented a system that automates class diagram generation from free-text requirement documents.
* The approach firstly applies NLP techniques to understand written requirements and then uses domain ontology to improve the performance of class identification.
* This methodology extracts candidate classes using part of speech tagger, a link grammar parser, parallel structure and linguistic patterns.
* Mich L. proposed a natural language processing tool named LOLITA (Large-scale Object-based Language Interactor, translator and Analyser).
* It is used to pre-process the user requirements; it includes all the tasks for natural language analysis.
* It’s built around SemNet which is a semantic graph that contains a large number of object and event nodes used to bridge the gap between object diagrams and requirements.
* This approach considers nouns as objects and use links to find relationships between objects. LOLITA extract only objects from natural language and it cannot distinguish between classes, attributes and objects.

**GATE Overview:**

* GATE is an open source framework developed using the Java programming language. It is used for developing software components that process natural language. GATE offers the foundational building blocks for higher level text understanding applications.
* It has an information extraction (IE) system called ANNIE (A Nearly-New Information Extraction System) which contains several language processing:
  + Sentence Splitter
  + Tokenizer
  + Parts of Speech Tagger
  + Syntactic Parser

**This paper also describes the following rules for extraction of classes for the class diagram:**

Rule 1: “All nouns are converted to entity types”.

Rule 2: “A gerund may indicate an entity type which is converted from a relationship type”.

Rule 3: a specialization’s relationship between entities: sentence’s structure “is a” can relate two nouns A and B to one another.

Rule 4: A noun such as “database”, “record”, “system”, “company”, “information”, “organization” and “detail” may not be considered as a relevant candidate for an entity type since it shows the business environment and logically have to be not included in the entity’s category.

Rule 5: every proper noun (Person name, Location name …) is ignored to be a class.

Rule 6: A noun such as “vehicle\_number”, “group\_no”, “person\_id” and “room\_type” may refer to an attribute type.

Rule 7: The genitive case, also called possessive case, often shows ownership. Hence, it can be used to extract attributes.

Rule 8: If consecutive nouns are present, check the last noun. If it is not one of the words in set S where S = {number, no, code, date, type, volume, birth, id, address, name}, most likely it is an entity type. Else it may indicate an attribute type.

[9] Rule 9: A noun phrase succeeding the “has/have” verb phrase may indicate the presence of attribute types.

**For associations’ extraction, the following heuristics are applied:**

Rule 10: A transitive verb can be a candidate for relationship type.

Rule 11: A verb followed by a preposition such as “in”, “on”, ‘to” and “by” can indicate a relationship type”.

Rule 12: “if a verb is equal to one of the following list {“include”, “involve”, “consists of”, contain, “comprise”, “divided to”, “embrace”}, therefore, this relationship can be aggregation or composition”.