# Hybrid Method for Tagging Arabic Text

**Presented By Ahmed Bukhamsin**

## Summary:

There are several important approaches to tagging like Hidden Markov Models and Finite State Transducers. Theses statistical approaches have some Drawbacks. They are inflexible and Based on small amount of information. There are some approaches based on the position of the word in the sentence but are not appropriate for tagging Arabic words because: Arabic has a weak positional constraint, the ambiguity in Arabic is enormous at every level and the absence of the short vowels increase the ambiguity. There are many methods of POS tagging which can be classified in three groups: 1. Linguistic approach 2. Statistical approach 3. Machine learning based approach. Hybrid method Combining more than one method so it get the advantages of each one of them. There are several signs in Arabic language that help to classify Arabic words like: affixes signs, grammar signs and pattern signs. Memory based learning is simple learning technique in which examples are massively retained in memory. Memory based learning system Contains two components: 1. A learning component which is memory storage 2. A performance component that does similarity-based classification. Uses of memory-based learning allows for easy integration of different information sources and can handle exceptions efficiently and has a number of advantages over statistical POS tagger. Rule based method gave 85% of correct result. The Hyper method gave 98.2% of correct results.

## References:

Hybrid Method for Tagging Arabic Text by Yamina Tlili-Guiassa

http://www.scipub.org/fulltext/jcs/jcs23245-248.pdf

Tagging by combining Rules based and memory based learning by Yamina Tlili-Guiassa and Laskri Mohammad Tyeb

http://www.ansijournals.com/itj/2006/679-684.pdf

## Obstacles:

It was default for me to understand the technique described in the paper since it is new for me and there are no enough details. So I tried to look for more details in the internet and read more.

The paper language was not clear in some part of the paper which made it more default for me to understand.

The result was not described properly in the paper but after do some searching I found similar paper for the same author with more details in the results.

Things I have learned:

I have learned that 2 days are not enough for preparing good presentation.

## Three true/ false questions :

Q1. Memory based approach is a hybrid method which combine rule based and statistical based approach. False

The correct answer is: Memory based tagging is a hybrid method which combine rule based and machine learning based approach.

Q2. Memory based learning system Contains two components A learning component which is memory storage and a performance component that does similarity-based classification. True

Q3. The memory based tagging system predict the tag using the similarity-based classification then verify the result using the rules. False

The correct answer is: The memory based tagging system predict the tag using the rules then the result using -based classification.