**APPENDIX A**

**POS TAGGING**

Part-of-speech tagging (POS tagging or POST), also called [grammatical](http://en.wikipedia.org/wiki/Grammar)tagging or [word-category](http://en.wikipedia.org/wiki/Lexical_category)disambiguation, is the process of marking up the words in a text (corpus) as corresponding to a particular [part of speech](http://en.wikipedia.org/wiki/Parts_of_speech), based on both its definition, as well as its context —i.e. [relationship with adjacent and related words](http://en.wikipedia.org/wiki/Lexicography) in a [phrase](http://en.wikipedia.org/wiki/Phrase), [sentence](http://en.wikipedia.org/wiki/Sentence_(linguistics)), or [paragraph](http://en.wikipedia.org/wiki/Paragraph). A simplified form of this is commonly taught to school-age children, in the identification of words as [nouns](http://en.wikipedia.org/wiki/Noun), [verbs](http://en.wikipedia.org/wiki/Verb), [adjectives](http://en.wikipedia.org/wiki/Adjective), [adverbs](http://en.wikipedia.org/wiki/Adverb), etc.

Once performed by hand, POS tagging is now done in the context of [computational linguistics](http://en.wikipedia.org/wiki/Computational_linguistics), using [algorithms](http://en.wikipedia.org/wiki/Algorithms) which associate discrete terms, as well as hidden parts of speech, in accordance with a set of descriptive tags.

**NLProcessor - Text Analysis Toolkit**

NLProcessor by Infogistics is a successor for a set of Natural Language Processing technologies developed in the 1990s at the University of Edinburgh. NLProcessor is an engine which handles so-called "low-level" text processing routines: tokenisation, capitalised word normalisation, sentence segmentation, part-of-speech tagging and syntactic chunking which are necessary steps in building many kinds of text handling applications.

NLProcessor outputs linguistic information by directly marking text with XML tags:

* Tokens are represented as "W" elements,
* Word-class [Part-Of-Speech information](http://www.infogistics.com/tagset.html) is provided in their "C" attribute,
* Noun and Verb groups are marked as NounGroup and VerbGroup elements, and
* Sentences are marked with "S" elements.

For example,

Consider the following sentence ie. “john has been given 25 bricks” with POS tags:

<S>  
<NounGroup>

<WC=NNP>John</W>

</NounGroup>   
<VerbGroup>   
<WC=VBZ>has</W>

<WC=VBN>been</W>

<WC=VBD>given</W>   
</VerbGroup>   
<NounGroup>

<WC=CD>25</W>

<WC=NNS>bricks</W>   
</NounGroup>   
<WC=".">.</W>  
</S>