Components

The general components of a typical (mostly factoid based) Question Answering system that we will also include are the following:

1. Question Processing

Under the Question Processing lies two main processes: Question Classification and Query Generation.

Question Classification

Question and answer types are categories in taxonomies (Moldovan et al. 1999) and defines a two-level hierarchy of questions. The simplest method to classify questions is the use of a set of rules that map patterns of questions into question types. The patterns are expressed by means of regular expressions on the surface form. The identification of the answer type is usually performed by analyzing the interrogative terms of the question (Handbook of Natural Language Processing). Question Classification is typically done by using a set of patterns, a rule based approach. Another method of question classification is through a statistical based approach using machine learning.

Query Generation

Once the question has been classified, the next process will be to select which of the words or entities will be used to retrieve the correct passage for the answer. This is accomplished by typically using keyword selection or again using a set of patterns.

2. Document or Passage Retrieval

Document or Passage Retrieval tasks are generally done through Information Retrieval systems rather than Natural Language Processing techniques as according to Gaizauskas (2004), using IR-Systems significantly reduces the time to search an entire document for the answer. When IR-Systems are used, it is done typically with ranking methods.

When using NLP techniques for Document or Passage Retrieval, passage retrieval is done through indexes. An example would be Litkowski (2000) who indexes with word pairs.

3. Answer Extraction

This component will complete the process of passage retrieval by corroborating the correct answer from the passages by either ranking them through the similarity in the text, ranking them based on the frequency of the words, ranking them according to rules already set in place, or ranking them through patterns.