

Reading comprehension activity #4: "What is Stemming?"

What is stemming?

Stemming is a linguistic process that consists of eliminating suffixes and prefixes from words to reduce them to their root or stem. In general, it is used to improve natural language processing, that is, the way computer programs can understand text.

Thus, by eliminating suffixes or prefixes, words of the same general topic, such as: cyclist and cycling, are changed to “cicl”, being what is known as “stem”, making it easier for programs to understand what the text is about.

Therefore, Stemming is a relatively basic technique and commonly used in combination with other natural language processing or NLP techniques.

What is stemming used for?

Stemming can be used for different tasks and, therefore, brings various benefits. Generally, it is used to reduce the complexity of a text, making it easier for software to process and understand the patterns of a topic more clearly.

Additionally, stemming can reduce the number of unique words while maintaining the meaning of most terms. Thus, by reducing words to their stem, it is easier to find patterns in a text, since words on similar topics will often become identical terms.

For example, if words like: “write” and “writer” are shown in various paragraphs, reducing them to their stem: “write” will cause fewer unique words to be processed and, at the same time, it will be understood that these terms belong to a group. same topic. As we have just seen, this makes it easier to detect patterns in content and understand fragments that deal with similar topics.

Differences between stemming and Lemmatization

Lemmatization is a term that is frequently heard when talking about stemming and, although they are very similar methods, there are important differences between the two.

On the one hand, stemming eliminates suffixes and prefixes from words, leaving a stem such as “cicl”, instead of “cyclist”, or “escri”, instead of “write”.

On the other hand, lemmatization uses dictionaries to find the base of the word, which is also known as the lemma of inflected forms. According to what has been said, the result would be that words like cyclist and cycling are shortened to, for example, “cycle” (its base without conjugations or paradigms) and, although this method is more complex and requires greater computing resources, it can offer better results, because all variations of a word are grouped into a base term.

Importance for SEO and marketing

Natural language processing (NLP) is an important part of search algorithms, so understanding the different methods that can help marketers can improve understanding of how these algorithms

work. NLP is used in different facets of marketing and is frequently adopted by various technologies.

Among its various uses, stem can be used in the creation of internal databases to improve the search functions of a website, in chatbots and in other external or internal features.

Taken from: <https://www.seobility.net/es/wiki/Stemming>