Language Models

Reference work entry

pp 1591–1594 | Cite this reference work entry

Encyclopedia of Database Systems

Djoerd Hiemstra

1177 Accesses   3 Altmetric

Synonyms

Generative models

Definition

A language model assigns a probability to a piece of unseen text, based on some training data. For example, a language model based on a big English newspaper archive is expected to assign a higher probability to "a bit of text" than to "aw pit tov tags," because the words in the former phrase (or word pairs or word triples if so-called *N-Gram Models* are used) occur more frequently in the data than the words in the latter phrase. For information retrieval, typical usage is to build a language model for each document. At

search time, the top ranked document is the one whose language model assigns the highest probability to the query.

Historical Background

The term *language models* originates from probabilistic models of language generation developed for automatic speech recognition systems in the early 1980s [9]. Speech recognition systems use a language model to complement the results of the *acoustic model* which models the relation between words (or...

ℹ This is a preview of subscription content, log in via an institution ↗ to check access.

Access this chapter

Log in via an institution

Subscribe and save

✔ **Springer+ Basic** €32.70 /Month

Get 10 units per month

Download Article/Chapter or eBook

1 Unit = 1 Article or 1 Chapter

Cancel anytime

Subscribe now →

Buy Now

⌃ **Chapter** EUR 29.95

Price includes VAT (France)

Available as PDF