

齐普夫定律

2021年10月29日

16:41

最省力法则：较好地解释了
齐普夫定律的内在成因和机制

齐普夫定律的基础

$$F_r \times r = C \text{ (C为常数)}$$

上式中： F_r 表示该词在文章中出现的频次， r 表示该词的等级序号。

◆齐普夫定律数学表示

- 将一篇较长文章(约5000字以上)中每个词出现的频次统计起来，按频次递减顺序排列。
- 并用自然数给这些词编上等级序号。
- 等级值和频次值的乘积是常数。

齐普夫定律的应用（仅适宜于中频词）

研究词频分布

1. 编制词表、制定标引规则、进行词汇分析与控制、分析作者著述特征

中频词往往是包含大量有检索意义的关键词。而一篇文献全文输入计算机后，计算机是很容易检出中频词的。

- 文献标引
- 情报检索
- 词表编制
- 科学评价

1. 关键词以精炼的语言反映了科研成果的精髓，可以展现学科领域的研究动向和发展动向
2. 基于网络环境进行的大规模词频统计分析，提高研究结论的可信度

网络信息计量学的应用：

■ 有学者选取一个月内某网站收到的页面请求进行分析，发现网页点击率和网页链接率都符合齐普夫定律，并认为可以通过齐普夫曲线分析网站的受欢迎程度。

■ 还有学者发现，二重齐普夫模型可用于挖掘网页访问量与目标网址数量之间的关系，表述网页的受欢迎程度。