基于《汉语主题词表》的自动标引测试与分析

常 春

博士 研究馆员

中国科学技术信息研究所

2019年10月18日

主要内容

一、《汉语主题词表》修订与重新编制进展

二、《汉语主题词表》服务系统自动标引功能

三、《汉语主题词表》自动标引基本思路

四、对《信息组织》一书进行自动标引的测试与分析

五、关于标引的遐想

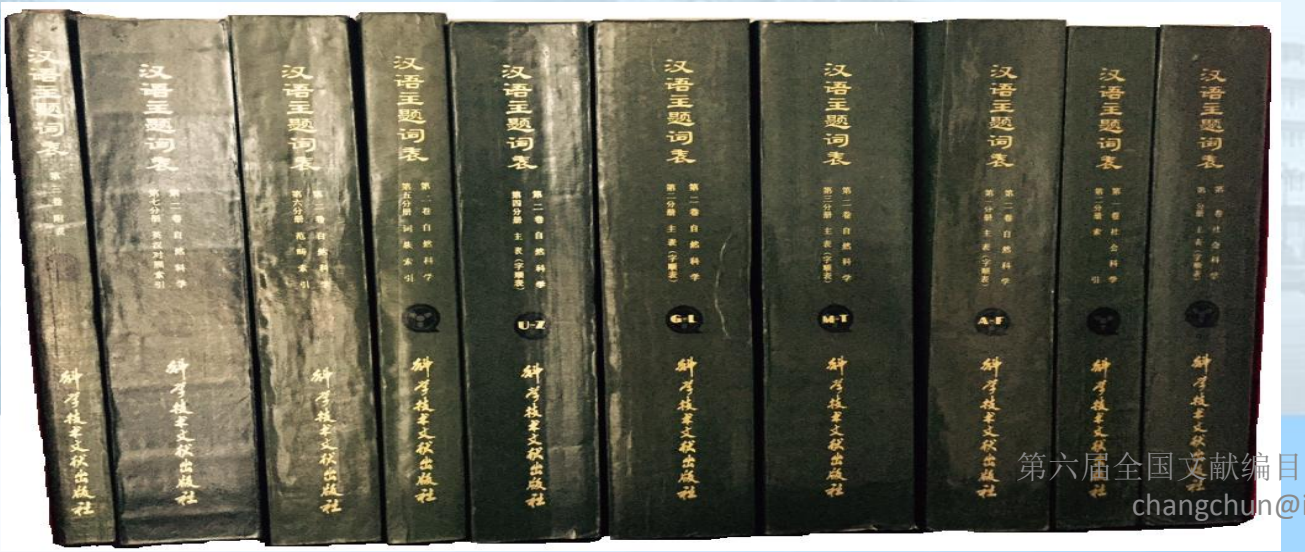
1980年版《汉表》——《汉语主题词表》

Concepts 优选词 91,158

Terms 总词汇 108,568

Social Sciences and Natural Sciences (including engineering technology)

社会科学、自然科学（含工程技术）



Volume 1 Social Sciences
第一卷　社会科学
First main list
第一分册　主表（字顺表）
Second index
第二分册　索引
The second volume of Natural Science
第二卷　自然科学
First to four main list
第一至四分册　主表（字顺表）
Fifth concept family index
第五分册　词族索引
Sixth category index
第六分册　范畴索引
Seventh English-Chinese index
第七分册　英汉对照索引
第三卷　附　表

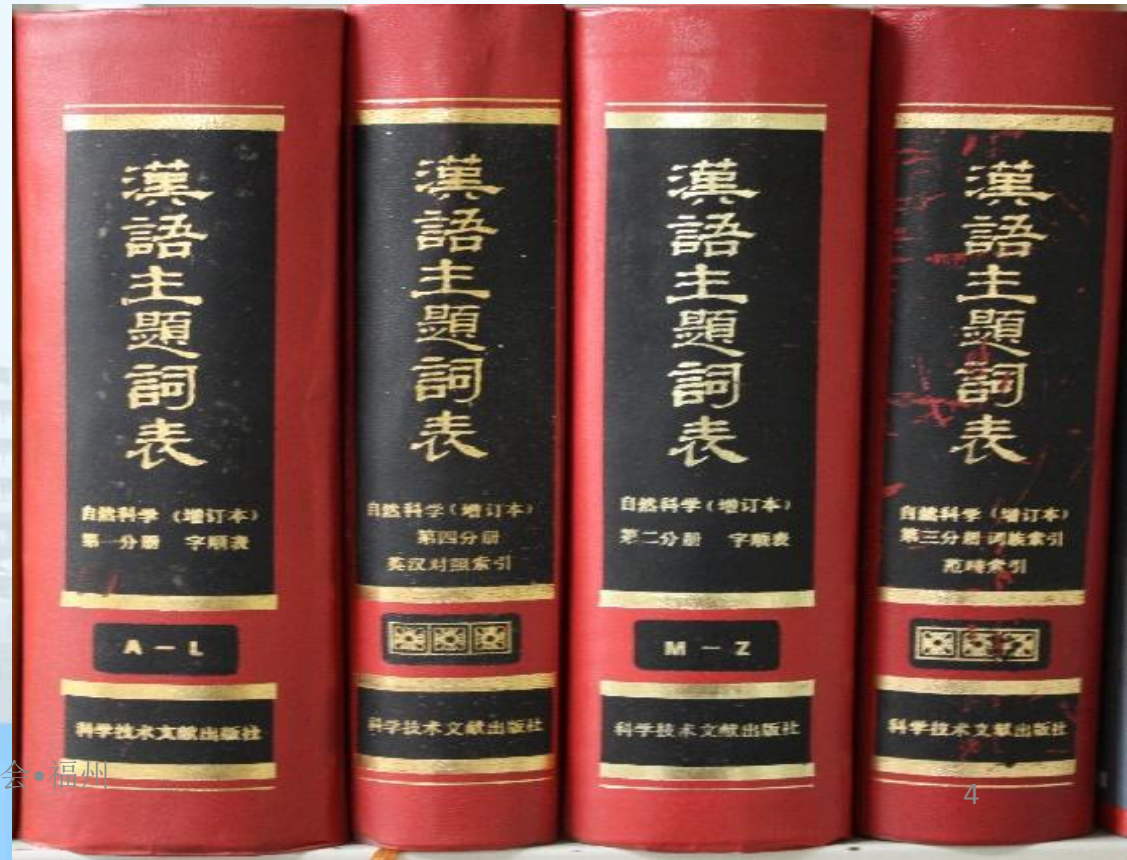
1991年版《汉表》

Including natural science, engineering technology, agriculture and medicine

包括自然科学、工程技术、农业、医学

Concepts 优选词 68,823
Terms 总词汇 81,198

First table A-L
第一分册 字顺表A-L
Second table M-Z
第二分册 字顺表M-Z
Third Index category
第三分册 词族索引 范畴索引
Fourth index
第四分册 英汉对照索引

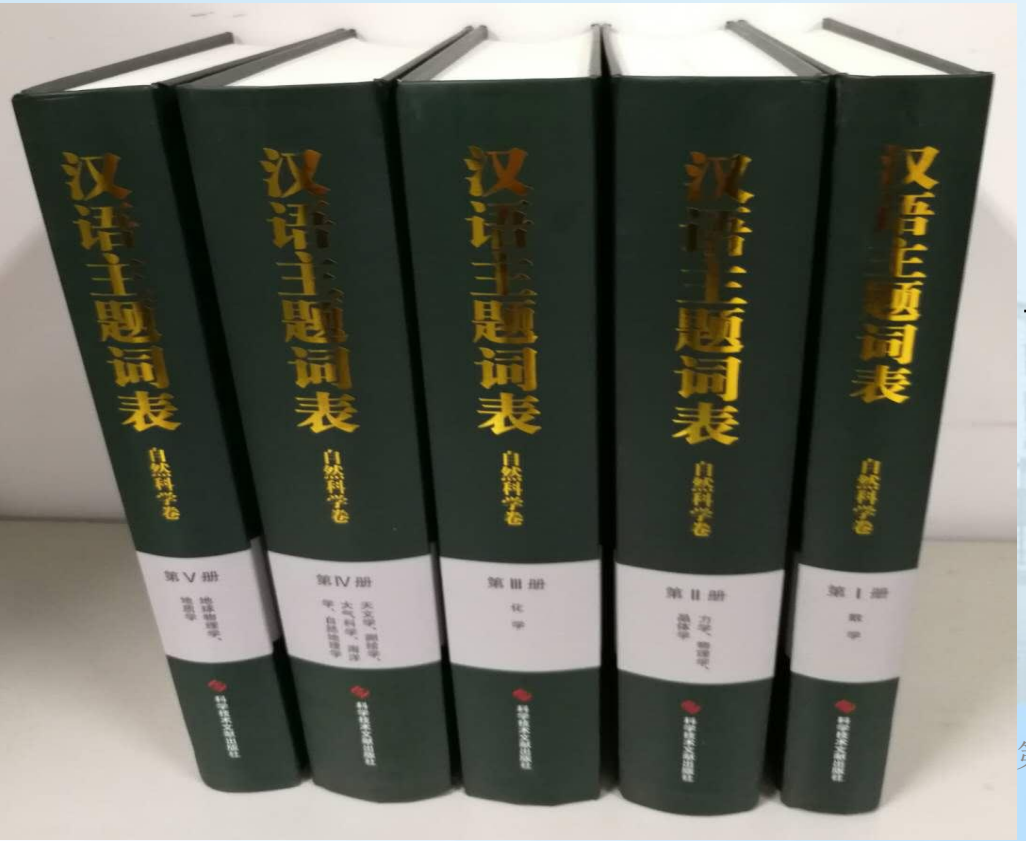


2014版《汉表》——《汉语主题词表（工程技术卷）》

只包括工程技术部分。2009年启动，2014年出版，共13册。大量增补词汇，使之更适用于计算机自动处理文献信息的需要，总词汇量36万条，优选词19.6万个，非优选词16.4万个。

分 册	词 量
第I册 工业基础科学与通用技术	28238
第II册 矿业、石油、天然气	30359
第III册 冶金与金属	45403
第IV册 机械、仪表	43468
第V册 动力与电工	33717
第VI册 武器、核技术、航空航天	30249
第VII册 电子与通信	36309
第VIII册 计算机与自动化	37579
第IX册 化工	32256
第X册 轻工	35597
第XI册 建筑与水利	44589
第XII册 交通运输	25813
第XIII册 环境与安全	23601

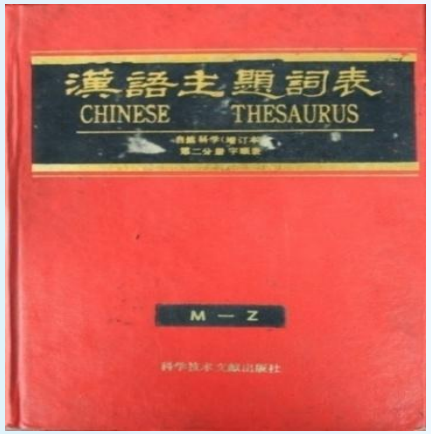
2018版《汉表》——《汉语主题词表（自然科学卷）》

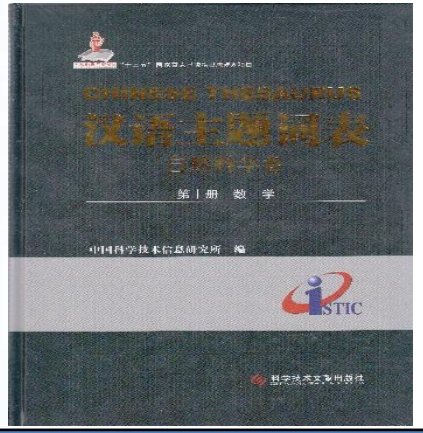


2015开始编制，2018年完成。6.5万条优选词，5.9万非优选词，总词汇量12.4万条。

Division 册	词量
Ⅰ 数学	18 586
Ⅱ 力学、物理学、晶体学	33 045
Ⅲ 化学	29 627
Ⅳ 天文学、测绘学、大气科学、海洋学、自然地理学	35 612
Ⅴ 地球物理学、地质学	21 389

Comparison of versions of Chinese Thesaurus《汉表》版本比较





	1980	1991	2014	2018
Domain 学科	全学科	自然科学增订本 （理工农医）	工程技术卷	自然科学卷 （缺生物科学）
Concepts 概念量	91,000	69,000	196,000	65,000
Terms 词汇量	109,000	81,000	360,000	124,000
Entrance rate 入口率	0.20	0.17	0.84	0.91

新型《汉表》建设进展

工程技术卷
（2009-2014）

Engineering
Technology
Volume

自然科学卷
（2015-2018）

Natural
Science
Volume

生物医学农业卷
（2019-2022）

Biology
Medicine
Agriculture

社会科学卷

Social
Science
Volume

项目申报工作

中华人民共和国科学技术部
Ministry of Science and Technology of the People's Republic of China

搜索

首页 组织机构 信息公开 科技政策 科技计划 政务服务 党建工作 公众参与 专题专栏

信息名称： 科技部关于发布科技基础资源调查专项2019年度项目指南的通知
索 引 号： 306-07-2019-680　　信息类别： 规范性文件2019
发布机构： 科技部　　发文日期： 2019年07月11日
文　　号： 国科发基〔2019〕236号　　效　　力：

科技部关于发布科技基础资源调查专项2019年度项目指南的通知

国科发基〔2019〕236号

二十六、汉语主题词表（生物医学农业卷）编研

工作内容：针对科技文献和科技基础资源中蕴含的相关科技知识、概念、术语等进行统一规范描述和知识关联的需要，选择生物、医学、农业三大领域，进行科技术语资源调查和新词发现，采集专业术语，对采集的术语进行领域范畴分类；识别同义术语，遴选叙词，建立资源描述术语规范及同义术语相互转换机制；建立中文、英文、拉丁文术语的映射关系；建立中文术语间知识的等级关系和相关关系；建立生物、医学、农业三大领域汉语叙词库（主题词表）及其服务平台；建立叙词库更新及维护的常规机制和程序。

http://www.most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2019/20190707/t20190712_147680.htm

主要内容

一、《汉语主题词表》修订与重新编制进展

二、《汉语主题词表》服务系统自动标引功能

三、《汉语主题词表》自动标引基本思路

四、对《信息组织》一书进行自动标引的测试与分析

五、关于标引的遐想

服务中心 我的图书馆 English http://www.istic.ac.cn/ 注 册 登 录 忘记密码

所馆主页

中国科学技术信息研究所
国家工程技术数字图书馆

国家工程技术数字图书馆

资源导航 | 发现服务 | 馆藏检索 | 科技信息 | 研究报告 | 科学评价 | 刊物出版 | 学术研究 | 院士著作馆 | 知识服务

服务平台 more

- 中国科技论文统计与分析网
- ISTIC专利信息检索与分析平台
- 国际科技创新与决策支持平台
- 中国科技情报网 www.chinainfo.gov.cn
- 全国科技查新网 chaxin.istic.ac.cn

科学评价 more

- 中国科技论文统计结果
- 中国科技论文统计源期刊

资源发现 Discovery 检索

馆藏资源 | 网络资源 | 特色服务 | 知识服务

- 外文学术期刊
- 外文学术会议
- 外文学位论文
- 国外科技报告
- 中文学术期刊
- 中文学术会议
- 中文学位论文
- 中国地方志
- 国家(地区)报告

科技信息

- 第四次中欧创新合作对话在布鲁塞尔召开 2019-04-17
- 新一代中继卫星天链二号01星成功发射 大幅提升我国数据中继卫星系统能力 2019-04-02
- 深度学习的下一站在哪里 2019-04-01

国家科技报告服务系统

《汉语主题词表》服务系统

所馆动态

- 关于举办2019年知识服务与情报工程学术交流会及情报工程前沿技术实践及应用
- 2019年全国第二期科技查新员培训通知
- 2019年全国竞争情报分析师培训班通知

>>更多

研究生培养

常春 changchun@istic.ac.cn

所馆主页

资源发现

全部 期刊 会议 学位论文 科技报告 专利 标准 图书 方志 法律法规

搜索 高级检索

资源及服务

服务项目 特色资源 专有系统

- 原文获取
- 收录引证
- 领域监测
- 专题服务
- 术语服务
- 培训讲座
- 在线咨询
- 知识服务
- 论文/专利分析
- 机构/团体合作

国家科技期刊开放平台

公告消息

更多 »

- 国庆节放假期间闭馆通知
- 关于与北京化工大学联合举办龙乐豪院士报告会的通知
- 中秋节放假期间闭馆通知
- 擦亮“中国名片” 吐尽一生芳华 - - 著名铁路专家、...
- 人工智能与人的智能 - - 李衍达院士应邀来院士著作...

资源通报

更多 »

- 资源介绍：日本产业时代社年鉴
- 资源介绍：ESDU系列
- 资源介绍：经济学人智库国家（地区）报告

友情链接

中华人民共和国科技部 国家科技图书文献中心 万方数据股份有限公司 科学技术文献出版社 中国科学院文献情报中心 机械工业信息研究院

冶金工业信息标准研究院 中国计量科学研究院文献馆 中国农科院农业信息研究所 中国医科院医学信息研究所 中国标准化研究院标准馆 中国化工信息中心


汉语主题词表服务系统 - Windows Internet Explorer
https://ct.istic.ac.cn/site/organize/welcome
百度搜索
收藏夹
建议网站
网页快讯库
汉语主题词表服务系统
页面(P)
安全(S)
工具(O)
中国科学技术信息研究所
INSTITUTE OF SCIENTIFIC AND TECHNICAL INFORMATION OF CHINA
《汉语主题词表》服务系统
开始体验
完成
Internet | 保护模式: 启用
85%
搜索一下
搜索
21:58
2019/4/18

服务系统


《汉语主题词表》服务系统
首页
术语服务
文本分词
自动标引
主题分析
学科分类
知识树DIY
术语服务
检索专业词汇语义信息。专业词汇486868条，关系654729个。更新词汇1455条。
进入
文本分词
提取文本中的专业词汇，对词汇表达进行规范控制。
进入
自动标引
中国科学技术信息研究所 | 关于汉表 | 版权声明 | 最新动态 | 服务办法 |
地址：北京市海淀区复兴路15号 邮编：100038 办公电话:010
国家工程技术数字图书馆
NATIONAL ENGINEERING TECHNOLOGY DIGITAL LIBRARY
个人登录
机构登录
账号密码登录
手机短信登录
用户名
用户名
密 码
登录密码
验证码
验证码
KDD7
登录
立即注册？ 忘记密码？

《汉语主题词表》服务系统 首页 术语服务 文本分词 自动标引 主题分析 学科分类 知识树DIY

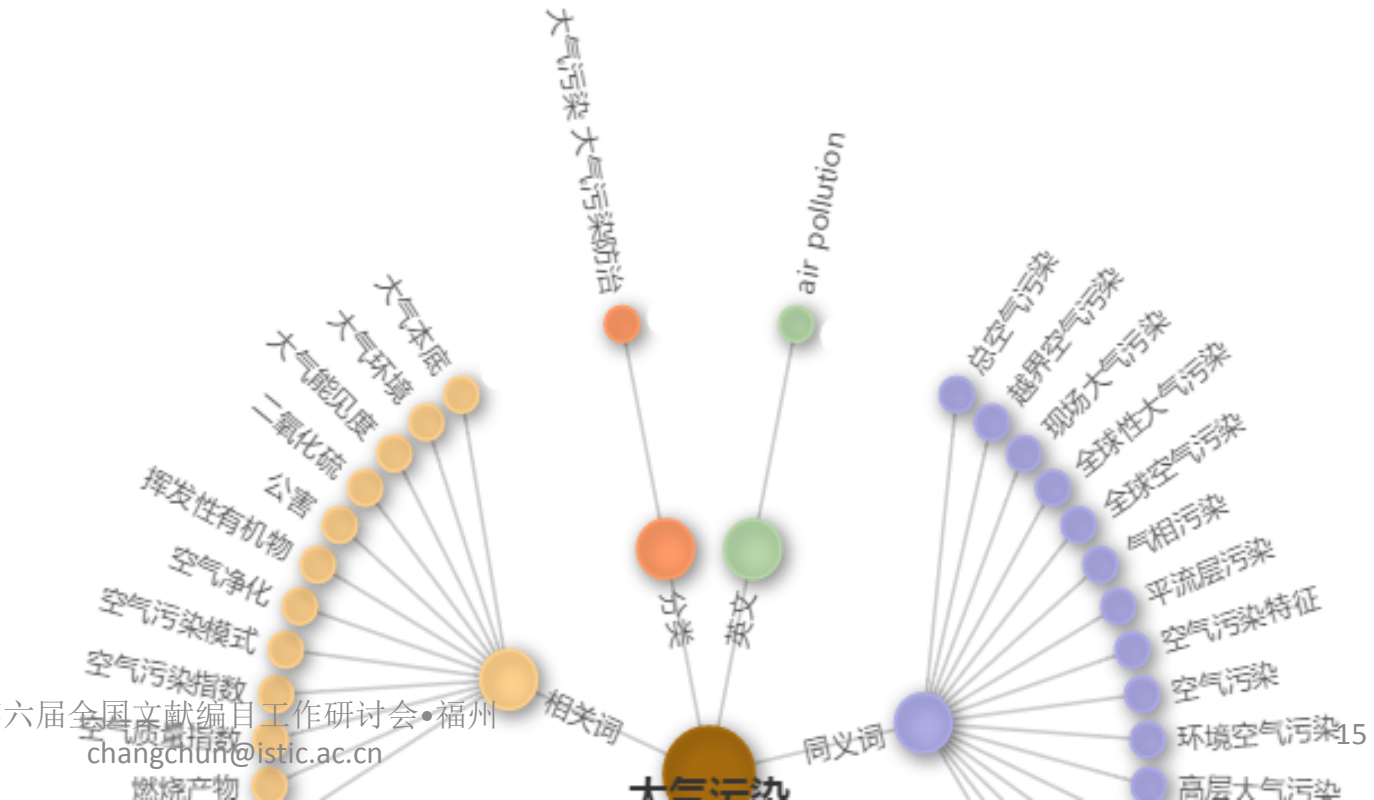
术语服务

术语检索 | 大气污染 搜索

请输入类目名称 确定

- A 马克思主义、列宁主义、毛泽东...
- B 哲学、宗教
- C 社会科学总论
- D 政治、法律
- E 军事
- F 经济
- G 文化、科学、教育、体育
- H 语言、文字
- I 文学
- J 艺术
- K 历史、地理
- N 自然科学总论
- O 数理科学、化学
- P 天文学、地球科学
- Q 生物科学
- R 医药卫生
- S 农业科学
- T 工业技术
- TB 工程技术（总论）

当前热点术语



术语服务

范畴树

- X 环境科学、安全科学
  - X5 环境污染、环境污染防治
    - X51 大气污染、大气污染防治

概念树

- 环境污染
  - 大气污染

大气污染

☆收藏 建议

概念属性 关注趋势 相关文献

来源	工程技术卷
分类	X51大气污染、大气污染防治
英文	air pollution
同义词	总空气污染、越界空气污染、现场大气污染、全球性大气污染、全球空气污染、气相污染、平流层污染、空气污染特征、空气污染、环境空气污染、高层大气污染、当地空气污染、大气污染特征、大气环境污染、本底空气污染、本底环境空气污染
上位词	环境污染
下位词	酸沉降污染、室内空气污染、燃煤大气污染、气体污染、颗粒物污染、光化学污染、废气污染、大气酸化、大气铅污染、大气氟污染、臭氧污染、城市空气污染、车内空气污染、P2O5大气污染

范畴树

- X 环境科学、安全科学
  - X5 环境污染、环境污染防治
    - X51 大气污染、大气污染防治

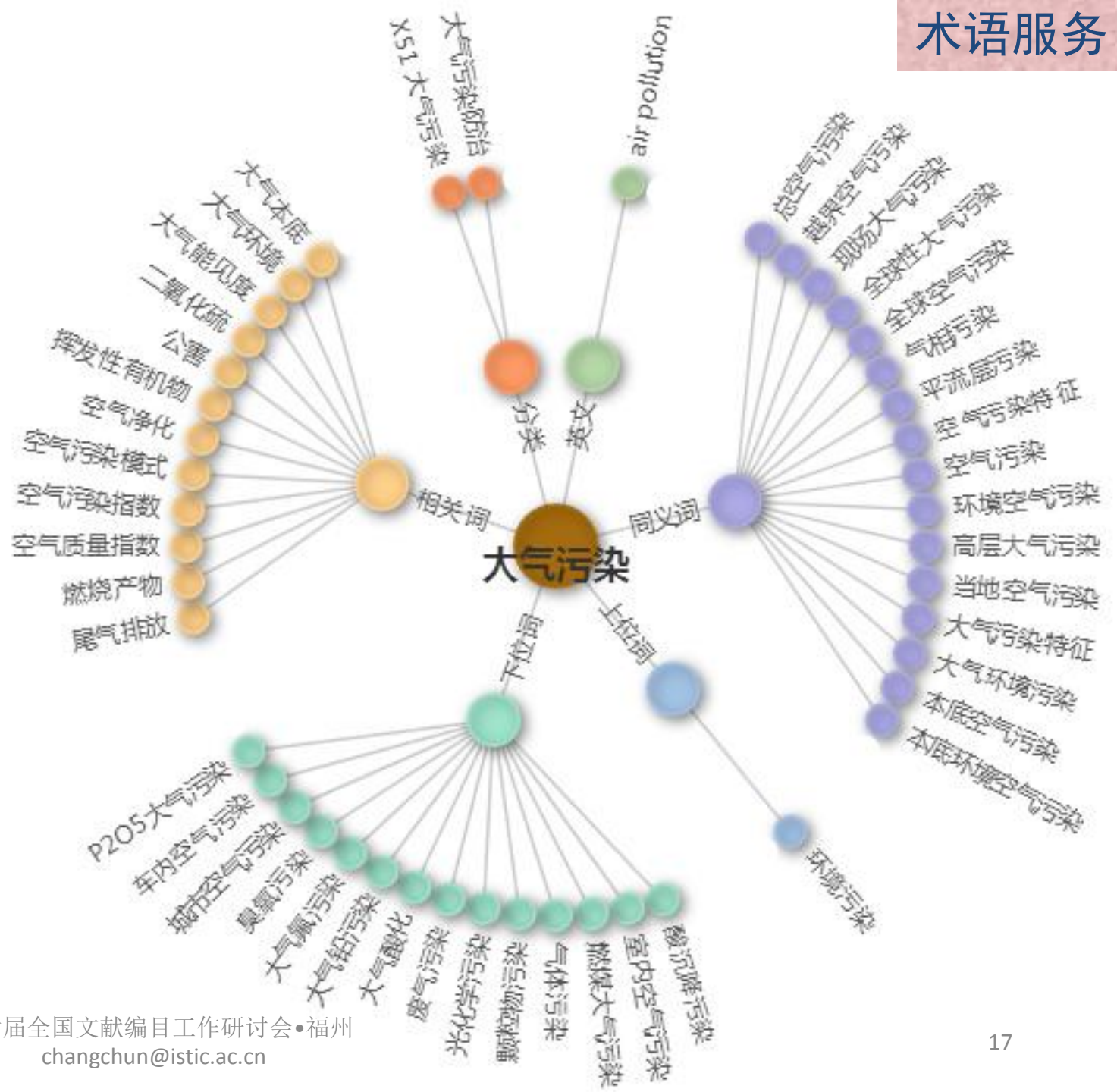
概念树

- 环境污染
  - 大气污染

大气污染

概念属性 关注

来源
分类
英文
同义词
上位词
下位词



基于生物种群增长规律的概念词频变化特征研究

常　春[1]，杨　婧[2]

（1.中国科学技术信息研究所，北京 100038；2.中国版本图书馆，北京 100005）

摘　要：【目的/意义】概念是知识的单元，在网络信息时代，如何鉴定概念的属性特征，是知识组织重要的研究方向。【方法/过程】基于生态学种群个体增长规律，对应文献数据库概念词频，逐年统计文献的数量，通过实证分析和总结概念从产生到成熟的生命周期过程。【结果/结论】知识组织体系的概念成熟过程，符合种群个体增长规律，概念词频达到最大以后，保持波动或有下降的趋势。

关键词：知识组织系统；生态学；概念词频；Logistic 生物种群增长规律；本体

中图分类号：G254.9　　DOI：10.13833/j.issn.1007-7634.2018.10.023

The Variation Characteristics of Concept Frequency Based on the Growth Rule of Biological Population

CHANG Chun[1]，*YANG Jing*[2]

(1. Institute of Scientific and Technical Information of China, Beijing 100038, China;

2. Chinese Version Library, Beijing 100005, China)

Abstract: 【Purpose/significance】Concept is the unit of knowledge. In the age of network, how to identify the attribute characteristics of concepts is an important research direction of the organization of information. 【Method/process】Based on the law of individual growth of ecological population, corresponded to the conceptual word frequency of the literature database, and counted the number of literature year by year; the paper analyzes and summarizes the life cycle for the concept mature process. 【Result/conclusion】The concept maturity process of knowledge organization system accords with the growth rule of individual population. After the maximum frequency of concept is reached, it keeps fluctuating or decreasing.

Keywords: knowledge organization system; ecology; concept frequency; Logistic growth rule of biological population; ontology

在自然界中，如果一个物种出现在一个可生存的新的自然环境中，该物种个体数量会表现出快速增长过程，增长到一定程度，由于生存空间逐渐饱和，可获取食物相对不足，物种个体数量会停止增长，甚至出现波动或下降，这就是生态学中有名的Logistic生物种群增长模型[1]。2016年，我们将生态学的原理和方法。应用到知识组织系统的构建和应用中，提出知识组织生态系统概念[2]，将物种与概念对应，环境与文献对应，研究概念遴选和概念关系的建立方法。基于生物种群增长规律，将单个生物种群个体数量的增长过程，与叙词表单个概念词汇在数据库中的文献数量增长过程对应，研究概念词频逐年变化规律，初步总结出概念词频的变化特征，并介绍了在概念术语遴选和概念关系建立等方面的应用[3]。本文使用全学科和领域数据进一步验证和完善概念成熟过程的研究，概念词频变化规律，并提出概念成熟的生命周期变化特征。

词频在情报学、自然语言学等研究领域有着广泛的应用。在知识管理中，马费成等基于词频统计研究知识管理研究热点[4]；陈果等将关键词词频在全局视角考察对领域研究特点的表征能力，通过关键词在领域内外的词频高低不同，提出领域度计算公式，揭示领域研究特点，克服了计量分析结果的主观性[5]。刘奕杉等以个人知识管理领域的文献为研究对象，通过热点聚类分析实证了高频词阈值鉴定方法的适用性[6]。通过词频统计，也可以进行共词、聚类等热点领域的研究分析，例如王宇等通过词频对国内自然语言处理研究现状进行的分析[7]。课题组在叙词表编制和应用中也大量进行了词频技术的应用[8]。总体而言，词频技术有广泛的研究和应用，但对应生态学物种个体角度的概念词频研究只有我们课题组进行过探索[9]。

收稿日期：2018-08-20

基金项目：国家重点研发计划（2017YFB1400200）；国家社会科学基金项目（15BTQ030）

作者简介：常　春（1966-），男，内蒙古人，博士，研究馆员，主要从事信息组织与数字图书馆研究。

·业务研究·
情报科学
第36卷 第10期 2018年10月

1　种群个体数量增长模型

研究生物种群个体数量在时间和空间上变动规律，在生物资源的合理利用、生物保护等方面具有重要的应用价值。生态学研究中，常常用数学模型模拟和研究种群变动规律。生态学中与种群个体密度有关的Logistic种群连续增长模型见图1[1]，在开始阶段（开始期），物种个体数量增长缓慢；随着物种个体的增加，物种个体密度进入迅速增加阶段（加速期）；当物种个体增加到环境容量的一半左右，物种个体增加速度达到最大（转折期）；以后随着资源和环境的限制，物种密度增加变缓（减速期）；当物种个体达到环境饱和数量时，物种密度停止增长（饱和期）。这是一个典型的"S"型增长曲线，对应的数学模型公式为：$dN/dt=rN(1-N/K)$，r为增长率，N为种群生物个体数量，K为环境可以容纳的最大种群生物个体数量。

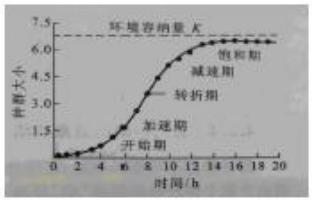
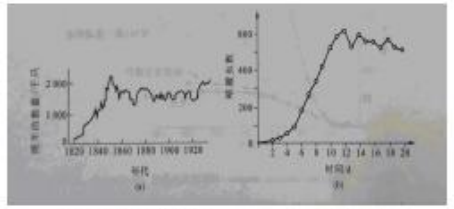


图1　种群在有限环境下的连续增长模型图

针对Logistic种群增长模型，也给出历史上绵羊和草履虫两个实际例子进行说明，见图2[1]。绵羊将近100年的种群增长过程，或者草履虫密度的变化过程，均表现出Logistic种群增长模型的"S"型曲线特征，并且补充说明，当环境发生波动时，种群数量也会发生波动，存在稍微超过种群密度平衡值的时期，是因为密度对增长率的作用存在一个时滞的原因造成的。



(a)塔斯马尼亚绵羊　　(b)草履虫

图2　所观察到的实际种群的个体数量增长

2　"本体"概念词频增长规律研究

2.1　假设的提出

依据知识组织生态系统研究框架[2]，将单个生物种群与单个概念对应，种群个体增长存在Logistic增长规律。假设单个概念所代表的文献数量随着时间的推移其增长也存在同样的规律，这样的假设是否成立，我们同样找一个例子来验证。在图书情报领域，国内从2000年左右开始，出现了"本体"概念，到2010年左右发展成知识组织的一个热点研究领域，以后"本体"概念研究逐步深入，保持着平稳的研究态势，用"本体"这个概念去验证是否存在这样的增长规律，文献数量通过以"年"为单位的概念词频来体现。

2.2　概念词频统计方法与变化分析

使用中国知网CNKI网络数据库，选高级检索，在检索框中输入"本体"进行主题检索，网站列出了历年文献总数，同时提供了文献数量按年统计发展趋势图。CNKI同时还提供了分领域（学科）进行检索的功能，"本体"涉及的学科包括：计算机软件及计算机应用、哲学、自动化技术、互联网技术、中国语言文字、电力工业、中国文学、有机化工、图书情报与数字图书馆、中等教育、美术书法雕塑与摄影、音乐舞蹈、戏剧电影与电视艺术、建筑科学与工程等学科。检索时，CNKI提示，只对前4万条记录进行了年度分组，"本体"全部文献记录显示为47966条，所以，完全可以用于统计对比词频大小、发展趋势等讨论。

首先统计"本体"在全部学科中历年的文献数量，结果见截图3。

图3　"本体"在全部学科中历年文献数量分布

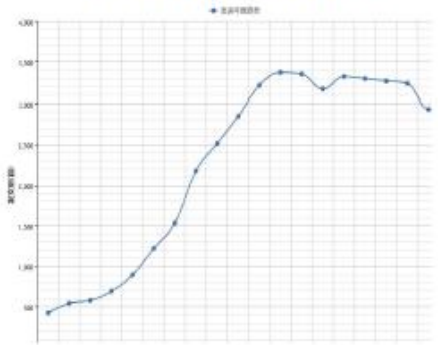


图4　"本体"在全部学科中历年文献数量变化趋势图

点击平台右下角提供的图表统计功能，生成"本体"概念词频发展变化统计图，见图4。可以看出，图4的概念词频增长趋势图与图1和图2总体趋势上存在完美的一致性，是一个典型的"S"型曲线图，同样存在一些波动，例如2012年是3165篇，比前后两年都低一些；2017年文献数降低，应该是

文本分词

文本

使用《汉语主题词表》中的规范术语和概念，在对文本准确分词与语义分析的基础上，自动为该文本标引恰当的主题词及学科分类。
注：作为功能演示，文本最长不能超过15000字符

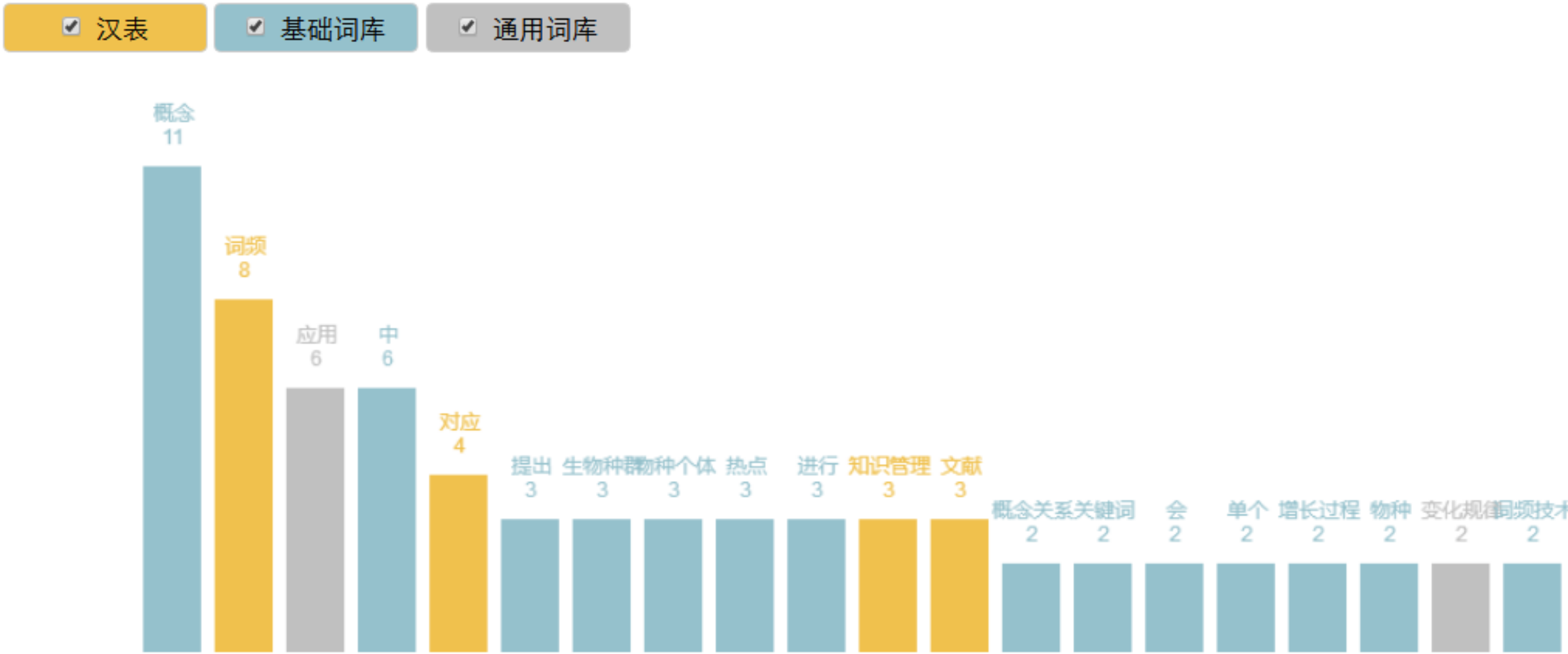
开始分析 清空

文本分词

汉表 基础词库

在 自然界 中 ， 如果 一个 物种 出现在 一个 可生存的 新的 自然环境 中 ， 该 物种个体 数量 会 表现出 快速增长 过程 ， 增长 到 一定程度 ， 由于 生存空间 逐渐 饱和 ， 可 获取 食物 相对 不足 ， 物种个体 数量 会 停止 增长 ， 甚至 出现 波动 或 下降 ， 这就是 生态学 中 有名的 Logistic 生物种群 增长模型 [1] 。 2016 年 ， 我们 将 生态学的 原理 和 方法 ， 应用到 知识 组织系统 的 构建 和 应用 中 ， 提出 知识组织 生态系统 概念 [2] ， 将 物种 与 概念 对应 ， 环境 与 文献 对应 ， 研究 概念 遴选 和 概念关系 的 建立 方法 。 基于 生物种群 增长规律 ， 将 单个 生物种群 个体数量 的 增长过程 ， 与 叙词表 单个 概念 词汇 在 数据库 中的 文献 数量 增长过程 对应 ， 研究 概念 词频 逐年 变化规律 ， 初步 总结 出 概念 词频 的 变化特征 ， 并 介绍 了 在 概念 术语 遴选 和 概念关系 建立 等 方面 的 应用 [3] 。 本文 使用 全 学科 和 领域 数据 进一步 验证 和 完善 概念 成熟过程 的 研究 ， 概念 词频 变化规律 ， 并 提出 概念 成熟的 生命周期 变化特征 。 词频 在 情报学 、 自然 语言学 等 研究领域 有着 广泛的 应用 。 在 知识管理 中 ， 马费成 等 基于 词频统计 研究 知识管理 研究 热点 [4] ； 陈果 等 将 关键词 词频 在 全局

词频统计


汉表
基础词库
通用词库
概念 11
词频 8
应用 6
中 6
对应 4
提出 3
生物种群 3
物种个体 3
热点 3
进行 3
知识管理 3
文献 3
概念关系 2
关键词 2
会 2
单个 2
增长过程 2
物种 2
变化规律 2
词频技术 2

《汉语主题词表》服务系统 首页 术语服务 文本分词 自动标引 主题分析 学科分类 知识树DIY changchun

标　题	基于生物种群增长规律的概念词频变化特征研究
摘　要	【目的/意义】概念是知识的单元，在网络信息时代，如何鉴定概念的属性特征，是知识组织重要的研究方向。【方法/过程】基于生态学种群个体增长规律，对应文献数据库概念词频，逐年统计文献的数量，通过实证分析和总结概念从产生到成熟的生命周期过程。【结果/结论】知识组织体系的概念成熟过程，符合种群个体增长规律，概念词频达到最大以后，保持波动或有下降的趋势。
关键词	知识组织系统；生态学；概念；Logistic生物种群增长规律
正　文	念词频的研究，可以对概念的成熟程度进行划分，也可以对概念关系的建立提供参考。 以上结论是在概念的分析基础上，通过部分概念词频的长期统计分析得到的结论，将来可以自动统计和测试某一数据库的全部概念，统计和验证分析其概念词频特征，以期有更多的发现。鉴于语言表达、文献主题、概念语义等的复杂性，本文结论仅限于统计概念的成立，更多概念的结果有待扩展不同数据库领域，统计和归纳更大数量的概念特征，在等同大数据的环境下归纳所有概念的成熟过程，有望对概念成熟规律进行进一步的修正。

开始标引　清空

主题词 Key words

学科分类 Classification

TP391	计算机信息处理
C8	统计学
X171.1	环境生态系统、生态环境
O212	数理统计
O212.1	一般数理统计

序号	主题词	文中词汇	相关度
1	自然语言处理	词频、自然语言处理	0.75
2	文献	文献	0.48
3	本体	本体	0.47
4	统计	统计	0.41
5	Logistic增长	Logistic增长	0.39
6	图书	图书	0.36
7	信息	情报	0.36
8	群体组合	种群	0.36
9	可视化	可视化	0.3
10	变化	变化特征、变动	0.3

主要内容

一、《汉语主题词表》修订与重新编制进展

二、《汉语主题词表》服务系统自动标引功能

三、《汉语主题词表》自动标引基本思路

四、对《信息组织》一书进行自动标引的测试与分析

五、关于标引的遐想

分类自动标引基本思路

1. 系统对文本进行分词、同义词归并、统计概念词频等自然语言处理工作，分词词库使用通用分词软件，加载《汉表》包含的术语，以《汉表》术语优先进行分词。
2. 按照标题正文等不同位置进行加权，统计文本分词结果，去掉停用词，将高频词进行排序，挑选 3 ~ 5 个《汉表》高频术语，提取其对应的分类号进行自动分析。
3. 如果《汉表》高频词具有的分类号是在同一领域不同等级的分类节点上，提取最专指的分类号即可。
4. 如果分类号处于同一领域，但处于上下等级关系的不同链中，则取两条链交叉的节点赋予分类号。
5. 如果处于不同的领域，则分别取最专指的分类号。

主题自动标引基本思路

1. 用户将文章标题、摘要、关键词和正文拷入相应的文本框。
2. 系统将所有文本优先基于《汉表》术语进行分词，以及基于《汉表》的基础词库进行分词。
3. 进行同义词归并，概念词频统计，并根据概念所处的不同位置对词频进行加权和排序。
4. 列出《汉表》的优选词以及相关度数值，用户可以根据自己标引的深度，确定相关度的阈值，系统就会给定相应数量的优选词标引。
5. 2019年上线的汉语主题词表服务系统包含了2014年完成的《汉表（工程技术卷）》和2018年完成的《汉表（自然科学卷）》，自动标引主要覆盖这两个大领域。

国家工程技术数字图书馆 NATIONAL ENGINEERING TECHNOLOGY DIGITAL LIBRARY

首页 资源导航 数据库 院士著作馆 服务项目 特色资源 专有系统 中信所 登录 注册

资源发现 Discovery

全部 期刊 会议 学位论文 科技报告 专利 标准 图书 方志 法律法规

航空母舰 搜索 在结果中搜索 高级检索

检索条件: 全部 ="航空母舰"

汉语主题词表辅助 执行

- 同义词
  - 航母
- 族首词
- 上位词
  - 水面战斗舰艇
- 下位词
  - 常规动力航母
  - 超级航母
  - 反潜航母
  - 核动力航母
  - 护航航母
  - 轻型航母

文献类型 执行

期刊(1,251)

全选 中文（共1415篇） 外文（共0篇） 每页显示10条 导出

排序： 相关度 更新时间升序 更新时间降序 出版时间升序 出版时间降序

1. 航空母舰

[图书] 董彩虹 ISBN:978-7-81059-139-3 北京:中国人民公安大学出版社 1998年 共168页

2. 航空母舰

[图书] 唐志拔著 ISBN:978-7-5065-3615-8 北京:解放军出版社 2000年 共122页

3. 航空母舰

[专利] 刘华友 CN200810068668.4 2008.03.21 共8页

摘要：现有航空母舰上存在着飞机起飞比较困难和降落比较危险的不足，本发明解决其不足所采取的技术方案是：在由舰体和飞行甲板等构成的航空母舰上，其特征在于设置有V形的起飞甲板和倾斜的降落甲板，V形的起飞甲板在舰艏的上升段低于、短于V形的下降段，起... 展开

原文获取

4. 航空母舰不过时

更多视图 >>

研究趋势

相关热图

主要内容

一、《汉语主题词表》修订与重新编制进展

二、《汉语主题词表》服务系统自动标引功能

三、标引的定义及相关概念

四、对《信息组织》一书进行自动标引的测试与分析

五、关于标引的遐想

《信息组织》一书按章节分割方式

总7章=14篇文献进行自动标引
第一章概论，概括介绍信息组织的主要内容，约4600字。
第二章信息组织的原理与方法，约12800字。
第三章分类法，分类是信息组织的重要手段，约26400字，按2篇文献处理。
第四章主题法，主要介绍叙词表编制与应用的相关内容，约54000字，按4篇文献
第五章本体构建与转化，主要介绍本体的构建、转化和概念关系的建立方法，约21600字，按2篇文献处理。
第六章信息描述，主要介绍信息描述、信息识别等元数据相关内容，约18400字，按2篇文献处理。
第七章知识组织生态系统，主要介绍知识组织生态系统的研究成果和进展，约24500字，按2篇文献处理。

《信息组织》一书按章节分类自动标引结果

整体分类标引类号类名及频次
TP391 计算机信息处理9
G25 图书馆事业、信息事业5
[P935] 生物地理学2
G255 各类信息资源工作2
G307 技术标准研究2
O177 泛函分析2
P962 自然资源调查、自然资源分析2
TP18 人工智能理论2
X171.1 环境生态系统、生态环境2
ZT74* 空间、位置、方位2

整体分类标引类号类名及频次
G254.29 其他知识组织系统1
N960* 控制论1
O435.1 光的反射、光的折射1
P343.3 湖泊、水库1
TN911 通信理论1
TS941.7 服装1
TS210.2 原粮1
TV1 水利工程基础科学1
ZT4* 属性、性能1
ZT6* 体系、结构、组成1

《信息组织》一书前10个主题词列表

主题词	文中词汇	总词频	权重总和
文献	文献	8	3.87
用户	用户、使用者	8	3.73
信息	信息、消息、情报	7	3.84
术语	术语	6	3.13
图书	图书、书	6	3.17
应用	实际应用	6	3.29
自然语言处理	词频	5	2.72
计算机	计算机	4	1.66
示例	实例	4	2.13
本体	本体、本体论	3	0.74

自动标引与人工标引核心概念对比

主题词	14篇文献词频	书核心概念	书中词频
文献	8	概念	1531
用户	8	关系	1347
信息	7	优选词	1043
术语	6	分类	845
图书	6	信息	728
应用	6	叙词表	633
自然语言处理	5	检索	523
计算机	4	主题	455
示例	4	知识	393
本体	3	系统	362

第四章主题法第一部分自动标引主题词相关度排序

序号	主题词	文中词汇	相关度
1	应用	推广使用、典型应用、实际应用	0.75
2	文献	文献	0.51
3	主题检索	主题检索	0.46
4	术语	名词术语、术语	0.46
5	搜索	信息检索	0.41
6	信息	信息、资讯	0.37
7	用户	用户	0.36
8	马铃薯	土豆、马铃薯	0.36
9	自然科学	自然科学	0.35
10	特性	特征、特点	0.35
11	粒度	颗粒度	0.35
12	图书	图书、书	0.34
13	钟表	表	0.34
14	标志	标识	0.34
15	本体	本体	
16	标准化	规范化	0.34
17	搜索引擎	网络搜索引擎、搜索引擎	0.34
18	专业	专业	0.32
19	交通工具	交通工具	0.32
20	奶茶	奶茶	0.32
21	数据库	数据库	0.31
22	小麦	小麦	0.31
23	乳制品	奶制品	0.31
24	自行车	脚踏车、单车、自行车	0.31
25	语义	语义	0.31
26	完备	完全的	0.31
27	网络	网络	0.31
28	水资源	水资源	0.3
29	环境	环境	0.3

再如，马铃薯代替土豆，土豆用马铃薯，它们是同义关系或用代关系。用、代的代号用汉语拼音符号 Y、D 代替，英文号表述为土豆 Y 马铃薯，马铃薯 D 土豆。在国际标准中规定推荐了通用的国际符号，如马铃薯 = 土豆，土豆→马铃薯，虽实际应用还没有完全推广普及。

奶豆、奶茶、奶泡泡、奶豆腐等各种不同类型的奶制品。这里的奶制品就是主题在内蒙古特产范围内找奶制品就能得到以上这些商品，这是网络购物的主题法以直接找奶茶，限定内蒙古特产，结果是大量的奶茶粉，还有奶茶片、奶茶伴的奶茶是个主题词，能找到大量不同商家的奶茶粉就是主题法的应用。

分析与结论

一、自动分类标引的前2个类目，分别为TP391 计算机信息处理（9次），G25 图书馆事业、信息事业。经过人工判断，信息组织、叙词表等这些术语还不是《汉表》的术语，故没有生成对应的G254信息组织类号，但生成了上位标引的G25类号。

二、自动主题标引的优选词与全文高词频术语本应该一致，但实际情况是两类词基本没有重复，TOP10术语中仅有一个“信息”是重复的，其他术语没有重复。分析其原因是许多高词频的术语在目前的《汉表》数据中不存在，所以机器没有用这些高频词进行标引。

三、扩大自动标引词的数量范围，看到一些与信息组织无关的术语存在，例如马铃薯、服装、奶茶等，经查是由于使用这些术语进行举例说明，多次提到，词频高，机器就把这些词标引为主题词，可见，自动标引的结果还需要人工进行审核和评判。

主要内容

一、《汉语主题词表》修订与重新编制进展

二、《汉语主题词表》服务系统自动标引功能

三、标引的定义及相关概念

四、对《信息组织》一书进行自动标引的测试与分析

五、关于标引的遐想

标引相关核心词

（一）标引手段：标引、辅助标引、自动标引、机器标引、自动抽词标引、可视化标引、受控标引、组配标引、自动赋词标引、自动化标引、自动主题标引.....

（二）标引方式：主题标引、分类标引、关键词标引、文献标引、学科标引、专利标引、专家标引、自然语言标引.....

（三）标引工具：词表、词库、分类法、主题词表、叙词表、中国图书馆分类法、汉语主题词表、知识组织系统.....

（四）普通名词：标引词、关键词、同义词、近义词、相关词、主题词、自由词、自然语言、字符串、通用概念、参考文献、知识组织系统.....

（五）动名词型：标引信息获取、词表编制、编目、分词、分类、聚类分析、全文检索、信息组织、知识组织、著录、信息描述、信息检索.....

2015年国家社科基金年度项目立项名单(节选)

1	序号	课题名称	负责人	工作单位	所在省市	项目类别	预期成果	计划完成时间	学科	批准号
2381	2380	面向叙词表构建的知识组织生态系统研究	常春	中国科学技术信息研究所	机关	一般项目	研究报告	2018-12-31	图书馆、情报与文献学	15BTQ030
2382	2381	基于小数据的高校图书馆用户电子资源使用习惯研究	杨涛	华南师范大学	广东	一般项目	专题论文集	2018-8-30	图书馆、情报与文献学	15BTQ031
2383	2382	近六十年来中国古籍出版研究	王育红	南通大学	江苏	一般项目	专著	2018-12-30	图书馆、情报与文献学	15BTQ032
2384	2383	古籍修复技术的科学化管理研究	马文大	首都图书馆	北京	一般项目	专著 研究报告	2018-6-30	图书馆、情报与文献学	15BTQ033
2385	2384	滇缅抗战文献的收集整理与研究	魏国彬	保山学院	云南	一般项目	专著	2018-12-31	图书馆、情报与文献学	15BTQ034
2386	2385	伪满时期日本战争罪行文献史料整理与研究	高承龙	延边大学	吉林	一般项目	专著	2018-12-30	图书馆、情报与文献学	15BTQ035
2387	2386	藏文古籍纸张的鉴别与修复综合技术研究	索朗仁青	西藏大学	西藏	一般项目	研究报告	2017-12-30	图书馆、情报与文献学	15BTQ036
2388	2387	建川博物馆馆藏二战时期侵华日军日记的整理、翻译与研究	瞿沐学	西华大学	四川	一般项目	译著 研究报告	2018-12-31	图书馆、情报与文献学	15BTQ037

阶段性成果（核心期刊论文）

[1] 常春. 面向叙词表构建的知识组织生态系统研究. 图书情报工作, 2016, 60(15)

[2] 杨婧, 常春. 基于生态位法则的概念稳定性研究. 图书情报工作, 2016, 60(13)

[3] 杨婧, 常春. 基于Logistic种群增长规律的概念词频变化研究. 情报科学,2017,35(8)

[4] 李永泽, 常春. 基于生态学能量传递的词族层次结构研究. 情报杂志, 2017, 36(3)

[5] 李永泽, 常春. 基于生态学信息传递的叙词表相关关系分析研究. 图书情报工作, 2017,61(18)

[6] 常春, 等. 基于生物多样性的图书馆信息资源建设研究. 图书馆理论与实践,2017(11)

[7] 常春, 等. 基于生态学能量流动原理的图书馆知识传递研究. 图书馆理论与实践,2018(1)

[8] 李永泽, 常春. 基于生态学种间关系的叙词表相关关系分类研究. 图书情报工作, 2018, 62(8)

[9] 邢福元, 常春. 基于生态学视角的叙词表概念多样性研究. 情报杂志, 2018, 37(11)

[10]邢福元, 常春. 基于生态学视角的叙词表概念稳定性研究[J]. 情报杂志, 2019, 38(7)

[11]常春, 杨婧. 基于生物种群增长规律的概念词频变化特征研究. 情报科学,2018,36(10)

[12]常春, 杨婧, 李永泽. 知识组织生态系统构架形成与研究进展. 图书情报工作, 2018, 63(7)

阶段性成果（硕士论文）

[1]杨婧. 基于物种稳定性的叙词表概念更新维护研究. 中国科学技术信息研究所, 2017

[2]李永泽.基于生态学种间关系的叙词表相关关系分类研究.中国科学技术信息研究所, 2018

[3]邢福元.基于物种多样性的叙词表概念稳定性研究.中国科学技术信息研究所, 2019

阶段性成果（著作）

目 录

全国哲学社会科学工作办公室

National Office for Philosophy and Social Sciences

坚持正确导向　突出国家水准　注重科学管理　服务专家学者

2019年8月国家社科基金年度项目、青年项目和西部项目结项情况

2019年08月30日17:03　来源：全国哲学社会科学工作办公室

点击查看：2019年8月国家社科基金年度项目、青年项目和西部项目结项情况

2019年8月国家社科基金年度项目、青年项目和西部项目结项情况

2019年8月，我办共验收572个年度项目、青年项目和西部项目结项申请。经同行专家鉴定，439个项目予以结项；112个项目暂缓结项（参照鉴定意见修改后报我办复审或重新申请鉴定）；20个项目被终止；1个项目被撤项。现将结项情况公布如下：

批准号	项目名称	成果名称	负责人	工作单位	证书号
A	B	C	D	E	F
15BTQ030	面向叙词表构建的知识组织生态系统研究	知识组织生态系统研究	常春	中国科学技术信息研究所	20192952

关于标引的遐（瞎）想.....

隐性主题：标引需要将隐性主题显性化，从而帮助用户更好的进行信息检索。从生物学视角看，标引标识类似于物种的“表现型”或农作物的“苗”，不同的苗代表不同的果实。

马铃薯

花生

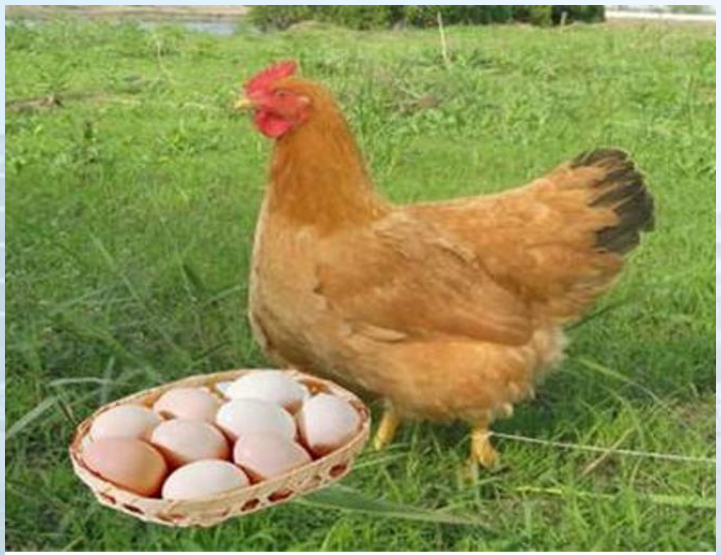
胡萝卜

关于标引的遐（瞎）想.....

关注性状：不同物种对人类来说关注的形状不一样，种植业关注植物的根茎叶花果实种子，养殖业关注家禽家畜的肉蛋奶形状，标引标注的主题是用户关注的主题，如果关注形状单一，就是单主题标引，相关工作例如重点标引、对口标引等。

生猪-瘦肉率

蛋鸡-产蛋率



奶牛-产奶量

关于标引的遐（瞎）想.....

全面标引：一个物种可能有多种价值，例如茴香，茎叶可食用，果实可以做食用香料、饲料添加剂等，也是重要的中药材。也即，一个物种需要标引多个主题，或者说对茴香来说，食用、香料用、药用都是人类关注的植物属性。推理为对用户有用的文献属性都需要标引，相关工作例如全面标引、多主题标引、分析标引等。

茴香–蔬菜

茴香–香料

茴香–药用

谢谢

changchun@istic.ac.cn