

计算机领域研究热点知识图谱分析

张小娟

(榆林学院信息工程学院, 陕西 榆林 719000)

摘要: 文章从文献计量分析的角度, 定量分析了计算机领域的2311995篇文章, 对国内计算机领域的主要研究机构、主要研究学者及该领域的研究热点行了统计与分析, 构建了国内计算机领域研究热点的知识图谱, 并对图谱进行了分析和解读, 以期获取我国计算机领域的研究热点有哪些, 有哪些主要的研究机构, 他们各自关心的研究主题又有哪些, 等等。

关键词: 知识图谱; 计算机; 研究热点; 数据分析

中图分类号: TP39 **文献标识码:** A

Knowledge Spectrum Analysis of the Hot Domains in Computer Science

ZHANG Xiaojuan

(School of Information Engineering, Yulin University, Yulin 719000, China)

Abstract: This paper quantitatively analyses 2311995 articles in the fields of computer science, from the view of literature metrology, collecting and examining data about major research institutions, main computer research scholars and the hot domains in this field, mapping knowledge research focus in the field in China, and analyzing and interpreting the spectrum to obtain what are the research hot domains, which are the main research institutions, and what their respective topics of research are concerned with, etc. in the field of Computer Science in China.

Keywords: knowledge spectrum; computer; hot domains; data analysis

1 引言(Introduction)

当今社会, 计算机的应用已渗透到社会的各个领域, 正在日益改变着传统的工作、学习和生活的方式, 推动着社会的科学计算, 成为国民经济基础性、先导性、战略性、支柱性产业, 是中国抢占国际经济制高点的重要引擎^[1]。但是计算机领域的学科发展现状如何, 该领域的研究热点有哪些目前还没有系统全面的研究。

中国知网是全球最大的知识门户网站, 集期刊、报纸、博士硕士学位论文、会议论文、图书、年鉴、多媒体教育教学素材为一体的知识服务网站。本文以中国最大中文期刊数据库——中国知网的海量数据为基础, 利用SPSS统计软件, 以及知识图谱的方法和工具对计算机领域的主要学术机构、学者、论文关键词等数据进行了统计分析, 以期获取我国计算机领域的研究热点有哪些, 有哪些主要的研究机构, 最主要的学者和他们的团队成员有哪些, 他们各自关心的研究主

题又有哪些等。

2 研究方法(Research method)

文献计量学方法是对文献和文献工作进行定量研究的方法^[2], 它采用数学语言进行定量分析, 以数学公式或图形来表现规律, 对问题的陈述及计算都采用简明的数学符号, 从而简化和加速了思维过程^[3], 具有深刻的描述性能和高度的概括能力^[4]。使用可视化的方式, 即用科学知识图谱的研究方法来定量, 并且形象直观地监测与评价学科的发展, 正是科学计量学、文献计量学、信息计量学等近10多年来一直共同关注的一个问题, 而且也已经取得了很大进展^[5]。

本文利用文献计量学及知识图谱的理念及技术对计算机领域的上百万篇文章进行定量统计分析。

3 研究步骤(Research steps)

3.1 数据来源

本文所统计分析的原始数据源自于中国知网数据库, 选

择对中国知网所收录的计算机相关领域2311995篇核心期刊文献进行统计分析，这些数据具有较高的准确性和权威性，是有价值的研究数据。

核心期刊是指某学科(或某领域)的核心期刊，是指那些发表该学科(或该领域)论文较多、使用率(含被引率、摘转率和流通率)较高、学术影响较大的期刊^[6]。由于核心期刊是期刊中学术水平较高的刊物，对核心期刊中收录的计算机领域研究文献进行检索分析，可以更准确的获得计算机领域的研究热点^[7]。

3.2 数据获取及预处理

(1) 数据检索

进入中国知网(CNKI)期刊数据库，在学科专辑导航中选择“计算机硬件技术”“计算机软件及计算机应用”“互联网技术”“自动化技术”“数据库”等计算机学科，共检索到2311995篇相关文献。

(2) 数据预处理

通过数据检索，检索出计算机领域的核心期刊文献之后，利用知网的文献题录导出功能，把文献的题名、作者、单位、关键词等题录数据导出，把所有的题录数据进行汇总^[8]，共得到2311995条题录数据，对导出的原始数据进行整合，便于之后对数据的分析。

3.3 数据统计与分析

3.3.1 单项关键词统计

对数据预处理后，通过对获取到的题录数据中关键词进行分列处理，其次对分列后的关键词进行整合，进而统计高频关键词^[9]。选取前几位的高频关键词包括“数据库、单片机、计算机、神经网络、遗传算法、数据挖掘、网络安全、仿真、PLC、数据采集、无线传感器网络、教学改革、故障诊断、云计算、图像处理、电子商务”等关键词。通过这些高频关键词，可以看出计算机领域的关注点。高频关键词分布见图1。

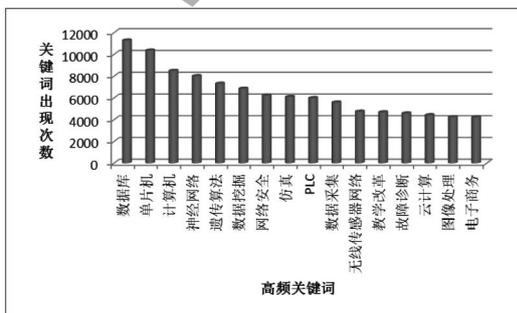


图1 高频关键词出现次数图

Fig.1 Frequency diagram of high frequency keyword appearance

3.3.2 关键词共现统计

共现分析是将各种信息载体中的共现信息量化的分析方法^[10]，可深刻、精确地挖掘隐含的或潜在的有用的文本知识，发现研究对象之间的亲疏关系，揭示研究对象所代表的学科或主体的结构与变化^[11]。

选取计算机相关学科所收录的2311995篇文献的前四个关键词进行关键词共现统计，构建关键词共现网络。通过关键词共现网络，可以发现，计算机领域的研究主题包括：计算机网络安全、计算机的应用、利用神经网络进行故障诊断、神经网络及遗传算法在数据挖掘中的应用、单片机与无线传感器、单片机与数据采集、基于云计算的数据挖掘、数据库与计算机网络等。关键词共现网络详见图2。

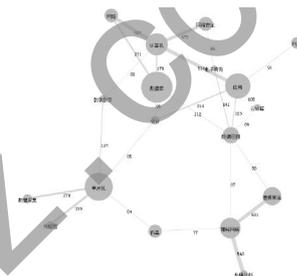


图2 关键词共现网络图

Fig.2 Key word concurrence network diagram

3.3.3 作者单位统计

对检索到的2311995条题录数据，利用Excel软件对作者单位进行分列汇总，按单位发表文献数量排序，发文篇数最多的为清华大学，其次为上海交通大学，排第三位的是浙江大学，发表文献频次从高到低前20位依次为：清华大学、上海交通大学、浙江大学、西北工业大学、华中科技大学、哈尔滨工业大学、国防科技大学、武汉大学、北京航空航天大学、东南大学、西安交通大学、华南理工大学、南京航空航天大学、同济大学、天津大学、重庆大学、中国科学技术大学、华中理工大学、东北大学、西安电子科技大学。发文数量居前20位的机构基本上为我国的重点本科院校。前二十名单位详见图3。

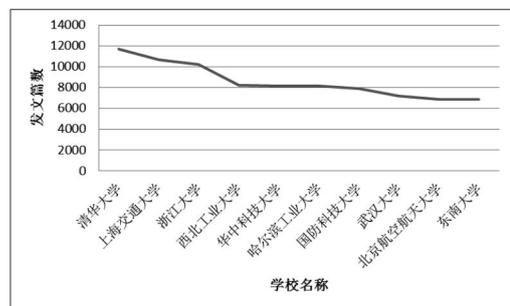


图3 机构发文篇数图

Fig.3 Number chart of institutional papers

3.3.4 作者统计

统计文献中作者出现频次可以分析得出计算机领域的领军人物^[12]。使用获得的2311995条题录数据,统计其中作者的出现频次,统计得到排序前五名的作者,依次为:王士同、杨静宇、于戈、周明全、赵海、王耀南、蔡自兴、陈志刚、桂卫华、潘泉、李仁发、杨炳儒、周明天、吴敏、王汝传。

这些高产作者均是计算机、信息技术领域的知名学者,排在前五位的作者简介:王士同,教授(博导),原江南大学信息学院院长,现任江南大学数媒学院首席教授;杨静宇,南京理工大学模式识别与智能系统国家重点学科学术带头人、博士生导师;于戈,现任东北大学计算机软件与理论研究所所长,2015年12月16日起任计算机科学与工程学院执行院长,一级教授,博士生导师;周明全,教授,博士生导师。北京师范大学信息科学与技术学院院长,教育部虚拟现实应用工程研究中心主任,教育部计算机科学与技术委员会软件工程教学指导委员会委员。赵海,现任东北大学计算机科学与工程学院计算机系统结构研究所教授、博士生导师,学科责任教授,面向先进装备制造业嵌入式技术实验室(省教育厅重点实验室)主任,享受国务院特殊津贴。相继创建了东大新业信息技术股份有限公司、沈阳市科技局无线传感器网络重点实验室、沈阳市发改委物联网应用基础研究工程实验室等。现任中国民主同盟中央委员、民盟辽宁省委副主委、辽宁省政协常委,曾任沈阳市人民政府参事、沈阳市委咨询委员会委员。曾获辽宁省第二批百万人才工程百人层次人选、辽宁省优秀博士学位论文指导教师、沈阳市青年专业技术人才、沈阳市优秀科技工作者称号。

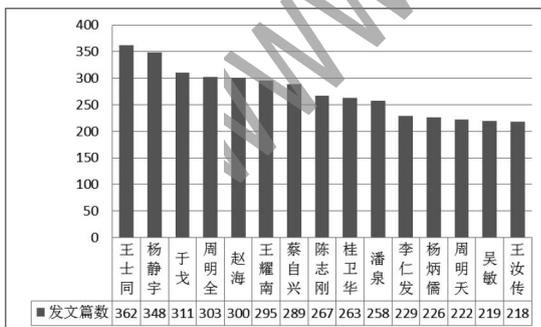


图4 高产作者发文数量图

Fig.4 The quantity chart of the prolific author

4 结论(Conclusion)

通过对计算机领域的2311995篇研究文献的文献计量学实证研究得出:计算机领域的主要研究机构集中于我国的重点本科院校,研究领域主要集中在“数据库、单片机、计算

机、神经网络、遗传算法、数据挖掘”等方面。随着人工智能技术的发展,与之相关的计算机网络安全、计算机的应用、利用神经网络进行故障诊断、神经网络及遗传算法在数据挖掘中的应用、单片机与无线传感器、单片机与数据采集、基于云计算的数据挖掘、数据库与计算机网络等主题成为计算机领域的关注热点。

参考文献(References)

- [1] Mohanapriya D.,Mahesh K..A novel foreground region analysis using NCP-DBP texture pattern for robust visual tracking[J].MULTIMEDIA TOOLS AND APPLICATIONS,2017,76(24):25731-25748.
- [2] Chen Wang,Hongxun Yao,Xiaoshuai Sun.Anomaly detection based on spatio-temporal sparse representation and visual attention analysis[J].Multimedia Tools and Applications,2017,76(5):6263-6279.
- [3] Luming Zhang,Yang Yang,Rongrong Ji,et al.Special issue on "visual semantic analysis with weak supervision"[J].Multimedia Systems,2017,23(1):1-3.
- [4] 吴运兵,阴爱英,林开标,等.基于多数据源的知识图谱构建方法研究[J].福州大学学报(自然科学版),2017,45(03):329-335.
- [5] 吴运兵,杨帆,赖国华,等.知识图谱学习和推理研究进展[J].小型微型计算机系统,2016,37(09):2007-2013.
- [6] 苏永浩,张驰,程文亮,等.CLEQS——基于知识图谱构建的跨语言实体查询系统[J].计算机应用,2016,36(S1):204-206;223.
- [7] 刘峤,李杨,段宏,等.知识图谱构建技术综述[J].计算机研究与发展,2016,53(03):582-600.
- [8] 刘知远,孙茂松,林衍凯,等.知识表示学习研究进展[J].计算机研究与发展,2016,53(02):247-261.
- [9] 王晓阳,郑晓庆,肖仰华.智慧搜索中的实体与关联关系建模与挖掘[J].通信学报,2015,36(12):17-27.
- [10] 杨良斌.数据挖掘领域研究现状与趋势的可视化分析[J].图书情报工作,2015,59(S2):142-147.
- [11] 王巍巍,王志刚,潘亮铭,等.双语影视知识图谱的构建研究[J].北京大学学报(自然科学版),2016,52(01):25-34.
- [12] 詹川.大数据研究的知识图谱分析[J].图书馆论坛,2015,35(04):84-91.

作者简介:

张小娟(1981-),女,硕士,讲师.研究领域:信息组织与信息检索,文献计量,知识图谱,数据分析。