

● 罗式胜

篇名关键词链特征的统计分析及应用

ABSTRACT Just like the fact of citing and being cited between documents, the using of literature keywords can integrate as an organic whole the documents that are in appearance irrelevant in and separated from one another, forming a group that is of a certain subject area. Using this method to study and predict the trend of development of a certain subject is more satisfactory and rational.

SUBJECT TERMS Title deywords — Characteristics. Literature — Keywords

CLASS NUMBER G256

象引证链一样,文献篇名关键词链也能把众多相关的文献或著者有机地联系起来。通过对各文献关键词链的定量统计能形成一种类似引证分析方法的关键词链分析。目前国内外关于关键词链分析的研究还不多见。本文试就此做一初步探讨。

1 文献篇名特征和篇名关键词链

众所周知,文献之间的引证与被引证特征能把表面上毫无联系的离散的著者或文献有机地联系起来,形成一个关于某主题领域的群体。事实上,利用各种文献的篇名特征也能把离散著者或文献有机地联结成一个某主题领域的有关联性的群体。

文献的篇名是该文献主要内容的概括。一般来说每篇文献都有一个或多个著者。大量的文献统计分析表明,各篇文献篇名特征的某些联系必然反映出各著者或文献之间的某些内在联系,这种联系为进行文献计量学研究提供了条件。

文献篇名的特征主要包括:

篇名关键词的概念及关键词的数量特征;

篇名主题词的概念及关键词的数量特征;

篇名关键词之间的搭配关系特征;

篇名主题词之间的搭配关系特征。

篇名是该文的窗口,通过它可以窥见论文内容概貌。

论文及其篇名都是由著者产生的,论文篇名在外部特征上的某些联系(例如有相同的关键词)必然反映出各著者在研究领域方向等方面的某些关系。对各篇名外部特征的这些联系,我们称之为“篇名关键词链”。这个“链”无形中把各离散的论文著者联系起来,形成以某专业主题为核心的一个个程度不同的著者群。这些著者群中,著者或许互不认识,但事实上,他们之间的的确确存在着潜在的联系。研究这种关系的形成、结构及其分布,为文献学的研究提供了新的素材。

那末,篇名关键词链是怎样确定的呢?具体地讲就是:如果两篇或多篇论文的篇名有一个或多个相同的关键词,则这两篇或多篇论文或著者之间则存在一种潜在关系,这种关系称为“篇名关键词链”。其中,相同关键词的数量决定着篇名关键词链的强度。如果两篇或多篇论文的篇名中有 P 个相同的关键词,那么这两篇或多篇论文或著者之间的篇名关键词链的强度为 P 级链引单位强度。

篇名关键词链作为一根根无形的“绳索”在客观上能起两个作用:一是把有关的著者链引成一个有机群体——篇名关键词链引著者群;二是把有关的文献链引成一个有机的群体——篇名关键词链引论文群。

如果某几篇论文篇名中至少有一个相同关键词,那么这几篇论文的著者形成一个群体,即篇名关键词链引著者群。一般情况下,如果这个“群”的有关论文中最多有 P 个相同的关键词,则称这个著者群是强度为 P 级的篇名关键词链引著者群。很明显,级数越高,篇名关键词链引著者群的相关性越大;反之则相关性越小。

实际统计时常会遇到这样一种情况:在一个 P 级篇名关键词链引著者群中,可能有 n 篇相关论文的著者是同一个人,若设这个著者为 M ,这时,我们可用一个权数来表示他,记作 M^n ,表示这个 P 级的篇名关键词链引著者群中著者 M 的权数为 n 。很明显,权数越大的著者,其在该篇名关键词链引著者群中越活跃。换句话说,权数越大的著者在进行该主题领域的研究中是多产著者。

与篇名关键词链引著群一样,同样可产生一个篇名关键词链引论文群。

即如果各论文篇名中至少有一个相同关键词,则这几篇论文形成一个群体,即篇名关键词链引论文群。

2 几点讨论

与引证研究分析一样,利用篇名关键词

链引,对文献及其著者的规律进行研究也是非常有实际意义的。但从分析论文篇名的关键词入手进行链引著者群研究还有许多应注意的地方:1. 对篇名关键词的分析可改为对篇名主题词的分析。这样做的优点是使自然语言规范化,使得对链引著者群的分析统计更精确、更全面,其缺点是比较麻烦。

2. 若把篇名关键词的分析推广到对正文或文摘关键词或主题词的统计分析,则效果会更佳。

3. 利用篇名关键词链对文献及著者群进行研究时,作为链扣的篇名关键词所涉及的主题专业范围越小越好,这样可以防止遗漏过多。例如,以“情报科学”这个关键词作为链扣所形成的篇名关键词链引著者群,往往会使许多属于“情报科学”这个主题专业范围的著者,出于其所著论文篇名中没有“情报科学”这个关键词而被排除在“情报科学”这个研究范围的著者群之外(如论文《齐普夫定律的来龙去脉》的著者)。如果把作为链扣的篇名关键词换为主题范围较小的“齐普夫定律”,则它所链引的有关齐普夫定律研究的著者群,相对来说遗漏的程度要小些。

4. 实际统计分析时,可以采用一些灵活的统计方法,例如可以把要求有两个相同关键词的条件改换成两个关键词虽不完全相同,但只要它们分别落入两个预先确定的关键词集中就行。

目前不少学者多通过对不同时期内学科文献数量的变化来研究和预测学科的发展、渗透、分化瓦解等趋势。这是一种传统的方法。这种方法并不完美,因为事实上不少著者可能是“多产”著者,某学科的文量虽大幅度增加,但进行该学科研究的人员阵容却没扩大,这时若断言该学科是处于发展阶段则是不全面的。而如果通过对某主题领域的研究人员的阵容的数量或质量的变化来预测,可能更科学些。

通过对各文献篇名关键词或主题词特征

的统计分析对文献作出许多有意义的揭示:

1. 统计各文献篇名中是否有相同关键词或主题词,可了解这些文献之间是否有联系,并能进而了解相应著者在研究领域的类同。

2. 能客观描绘出一个具有一定特定主题研究范围的著者群体的具体地区或部门的分布状况。

3. 能测试出著者群体相关联的强度。

4. 对某一主题范围的著者群按不同年代进行时间分段追踪比较统计,能从该著者群在著者姓名、著者数量、地区分布、部门分布的变化比较分析中,了解该主题研究的时间变迁趋势。

5. 分国别统计篇名中具有 n 个相同关键词或主题词的若干个著者群,能了解该主题研究领域在国际范围内的分布和发展趋势。

6. 国际范围内,文献篇名中具有相同关键词或主题词的著者可形成一个相关著者群。相同关键词或主题词数量越多,相关著者群的关联程度越强,这个关联程度强的著者群可称作该主题研究领域的核心著者群。这个核心著者群研究重点一致,是进行国际学术通讯交流,以文会友的较佳网络。

3 展 望

1. 篇名关键词链与文献引证链(引证分析)一样,是有实用研究价值和开发潜力的文献计量学研究课题,特别是如前所述,如果把对篇名关键词的统计分析推广到整个学科文献(文摘或正文)关键词(或主题词)的统计分析,则可使该项研究更趋于合理和完美。

2. 如果说文献引证链是通过文献的引证

或被引证特征开拓的对文献潜在关系的一种文献计量分析研究方法——引证分析,那么,文献关键词链特征将从另一个角度开拓对文献潜在关系的另一种文献计量分析方法——关键词链分析。两种方法将对充实文献计量学理论宝库起到积极作用。

3. 对关键词链的研究刚刚开始,研究领域十分广阔,还有许多课题值得研究,如关键词链的耦合、关键词链的网络结构等。

4. 论文书写的规范性对进行关键词链统计分析非常重要,如果能象正规的论文列出一条条参考文献一样,在该论文题目或文摘之后也正规列出该文的若干主要关键词,进行关键词链研究会更方便和规范。

5. 文献计量学中,任何一种分析方法(如最常用的引证分析)都不可能没有缺点,关键词链统计分析也一样,其统计分析的精确性常常受到论文著者在选择关键词时的人为因素影响。但只要能有所规范,这种影响会降到最低限度,关键词链引证统计分析仍不失其有较好的应用前景。

参考文献

- 1 王万宗. 情报学概论. 北京:北京大学出版社,1988
- 2 罗式胜. 著者链概念及其应用. 情报学报, 1984, (4)
- 3 严怡民. 情报学概论. 武汉:武汉大学出版社,1988

罗式胜 现为广州中山大学信息管理学副教授,系副主任。已发文 70 余篇,出版《文献计量学引论》等专著 5 种。通讯地址:广州市。邮编 510275。

(来稿时间:1994-02-01. 编发者:翟凤岐)