

石油勘探开发领域空间信息检索方法

王志宝¹, 王成波², 文必龙¹, 夏昊², 乔彦友²

1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318

2. 中国科学院 遥感与数字地球研究所, 北京 100101

摘要: 提出了面向石油勘探开发的领域空间信息检索技术方法, 给出了文档空间主题计算模型, 设计了支持文本与三维空间查询的高性能索引结构, 在原有研究的基础上实现了原型系统 PetroScope. 经过真实数据的性能测试和比较分析, 原型系统的准确度得到了一定程度的提高, 性能能够满足该领域空间信息检索的需求.

关键词: 空间信息检索; 石油勘探开发; 空间信息提取; 实体消歧; 混合索引结构

中图分类号: TP391; P208 **文献标识码:** A **文章编号:** 0455-2059(2017)05-0576-06

DOI: 10.13885/j.issn.0455-2059.2017.05.002

Spatial information retrieval method for petroleum exploration and production

Wang Zhi-bao¹, Wang Cheng-bo², Wen Bi-long¹, Xia Hao², Qiao Yan-you²

1. Computer & Information Technology College, Northeast Petroleum University, Daqing 163318, Heilongjiang, China

2. Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China

Abstract: A technical framework of spatial information retrieval was put forward for petroleum exploration and production, a method of spatial theme mining proposed and a high-performance hybrid index structure supporting query composed by text and 3 dimensions space were also designed. The prototype named PetroScope based on original research was implemented. It was proved that the PetroScope could meet requirements of spatial information retrieval in petroleum exploration and production with its accuracy and performance verification.

Key words: spatial information retrieval; petroleum exploration and production; spatial information extraction; entity disambiguation; hybrid index structure

石油工业是信息工业, 从石油勘探、油藏评价、油田开发到油气集输都依赖于对各类信息的采集、检索、处理和分析. 数字油田建设产生了巨量的数据资源, 这些数据很多以文本的形式存在, 蕴含了丰富的业务知识, 快速准确地信息检索是该领域重要的业务需求.

传统信息检索技术提高了文档查找效率, 却忽略了文本中的空间特征, 导致检索性能不佳. 石油勘探开发认识和研究的对象是地下空间内地质构造和油藏^[1-2], 石油勘探开发活动产生的数据几乎都具有显式或者隐式的空间参考. 参考地理信息检索框架, 利用自然语言处理技术提取空间参

收稿日期: 2015-05-12 **修回日期:** 2016-10-08

基金项目: 东北石油大学青年科学基金项目(NPUQN2014-18); 中国海洋石油总公司油气重大专项(2011ZX05023-005-012)

作者简介: 王志宝(1981-), 男, 黑龙江穆棱人, 副教授, 博士, e-mail: zhibao_wang@126.com, 研究方向为地理信息系统与信息检索.

考,计算文本的空间主题,研究空间信息检索方法,对提高石油勘探开发信息检索的性能具有现实意义.

1 相关工作

空间信息检索起源于互联网上的地理空间信息检索(geographic information retrieval, GIR), Web 用户提交的信息检索请求中,有超过 18.6% 的请求包含至少 1~2 个地理约束条件^[3]. 基于关键字的信息检索系统并没有考虑空间关系,此类查询系统的准确性普遍不高. GIR 研究工作由此开始,其方法是将文档中的地理空间信息提取、存储、索引和排序过程进行特殊处理. 与本研究相关的研究有两方面:地理信息抽取(使用自然语言处理和机器学习技术从文档或者查询中解析地理空间信息);混合索引结构(研制空间-文本混合索引,满足空间-文本混合查询的性能需求).

地理信息抽取:集中在地理命名实体识别、实体消歧、地理求焦等任务,目的是获取文档空间主题. 地理命名实体识别有机器学习,例如 SVM^[4]、CRF^[5-6];基于规则与统计结合的方法,例如规则与 CRF 相结合的方法^[7]、基于规则与 Bison 结合的方法^[8];基于知识的方法,例如使用地名词典^[9-10]或者地理本体^[11-12]等方法. 地理消歧主要去除地理命名实体的歧义,主要有地名/地名、地名/人名消歧,方法有基于地图的方式^[13-14]或者基于知识的方式^[15-16],也有用基于机器学习的方法,例如共现模型^[17];地理求焦是从文档的相关地名中综合文档空间范围,有基于认知地理参考树实现地理聚焦方法^[9,18]. 石油勘探开发领域主要关注地下空间的石油地质实体,具有不同的空间关系语义,基于地理知识的方法不能适应这种领域特殊性. 由于大规模领域语料库的缺失,基于统计的方法也需要进一步改进.

混合索引结构:包括文本与空间的松散组合方式^[19]、以文本索引为主的混合索引^[19-20]、以空间索引为主的混合索引^[19-21]、文本和空间方式融合索引^[22-23]. 第一种方式两种索引间没有或者只有松散的联系,需要分别查询两种索引,对结果进行交集运算. 第二种索引先以倒排文件建立主索引,然后每个关键字节点保存一个指向 R-tree 的指针,每个 R-tree 对包含这个关键字的集合进行空间索引. 第三种索引是以 R-tree 索引建立主索引,每个叶子节点指向其所覆盖的文档子集合的倒排文

件. 前三种方式容易实现、兼容性好,但查询很难支持文本和空间的同时剪枝处理,第四种是对以空间为主的混合索引改造,对节点结构、查询算法改进. 例如 Cong 等^[24]将 R-tree 中的非叶子节点视为虚拟文档,表示该子节点中所有项的文档集合,用来估计一个查询与该节点所有文档文本相关性的界限值,实现文本与空间的同时剪枝. 在石油勘探开发领域,空间查询是三维的,既有平面空间的区域查询,也有垂直空间的地层查询,虽然三维 R-tree 支持三维空间查询,但其 3 个坐标维度是同构并定量的,由于空间知识库的限制,可获取的垂直空间信息是定性的. 垂直空间查询的空间关系包括地层间的层级关系、邻接关系、包含关系,这些特征也是传统空间索引不能支持的,本研究将研究一种新的多维混合索引结构满足本领域文本和空间的“语义-定性-定量”混合查询.

2 方法

2.1 总体技术框架

针对领域的三维空间性和特定语义,设计了一个 5 阶段的空间信息检索框架:信息采集、信息处理、空间主题计算、建立多维索引和多维混合查询.

信息采集从石油企业的门户网站、数据库、文献资料库等异构跨系统的数据源中抽取大量领域文档,针对每类数据源开发一种爬行器,在前期研究中已实现可扩展的异构数据源爬行框架^[25]. 抓取文档集之后开始进入信息处理阶段,首先要对各种格式的文档进行归一化处理,转化为统一的内部格式,然后将文档进行分词处理,最后对文档执行词性标注、命名实体识别、依存句法分析基本的自然语言处理(natural language processing, NLP),为文档的空间主题计算做准备. 文档空间主题是文档的水平空间范围与垂向空间范围,首先识别出领域特有的井、构造、油藏、地层等诸多空间实体,然后确定空间实体所属的类型,确定是水平维度还是垂直维度的,最后综合所涉及的空间实体集合计算文档的水平兴趣区域(area of interest, AOI)与垂直兴趣层(stratum of interest, SOI). 为满足专业人员主题查询、空间查询和组合查询的需求,建立文档集的多维索引结构,优化使用文本倒排索引、R-tree 空间索引和以地层本体为基本结构的索引. 从这个过程中看,文档空间主题计算和多维索引结构是研究的重点,已有的 GIR 模

型和方法不能适应本领域的特质需求.

2.2 文档空间主题计算模型

文档空间主题计算目的是确定文档的空间范围 AOI 与 SOI. 文档空间化有 3 个步骤: 实体空间识别、空间实体消歧和空间聚焦.

2.2.1 实体空间识别

命名实体识别技术在 NLP 领域已经比较成熟, GIR 促进了地理空间信息抽取研究, 目前比较主流的方法有基于机器学习方法^[26-27]和基于知识的方法^[10-11, 28]. 本研究采用半监督学习与领域空间本体结合的方法. 石油勘探开发空间本体的 Abox 生成领域空间辞典, 利用辞典将词汇序列标注为空间实体. 由于没有标注的领域语料库, 但未标注语料却很丰富, 因此采用基于条件随机场的半监督学习^[29]的方法进行潜在的空间实体探测, 使用特征包括临近词、词性、词后缀、词频和浅层句法特征.

2.2.2 空间实体消歧

实体消歧目的在于明确空间实体的语义, 例如“萨尔图”既是地理空间实体“大庆市萨尔图区”, 也是地质空间实体“大庆油田萨尔图油层”, 由于地质实体的命名习惯, 这种情况在本领域大量存在. 石油勘探开发领域具有丰富的知识源, 包括业务模型、数据模型和行业标准, 这些知识源随企业的业务更新, 时效性很好. 从各类知识源中集成石油勘探开发空间本体, 利用它进行实体消歧. 实体消歧算法输入: 具有歧义的空间命名实体集合 $W\{w_1, w_2, \dots, w_n\}$, 空间命名实体在文本中的上下文集合 $C\{c_1, c_2, \dots, c_n\}$, 任意的 $c_i \in C$ 是一个词汇集合, 以及领域本体 O .

实体消歧算法步骤:

- 1) 针对有歧义的空间命名实体 w_i , 领域本体 O 为基础, 抽取每种可能解释的语义子图 $G_i\{\{n_1, n_2, \dots, n_n\}, \{r_1, r_2, \dots, r_n\}\}$, n 为节点, r 为关系;
- 2) 依次计算 w_i 的每个语义子图 G_i 与实体的上下文 c_i 的相似度 CM, c_i 也是图结构;
- 3) 取 CM 最大的 G_i 对应的概念实例作为 w_i 的意义;
- 4) 为下一个有歧义的空间命名实体 w_{i+1} 消歧, 直到最后一个.

CM 度量模型使用概念密度模型^[30]:

$$CM(m, f, n) = m^a \left(\frac{m}{n} \right)^{\ln}, \quad (1)$$

其中, m 为语义子图 G_i 与实体的上下文 c_i 的交集的

大小, n 为语义子图 G_i 的节点数量, f 为常见度排名, a 为约束因子, 取 0.1.

2.2.3 空间聚焦

经过实体识别和实体消歧的处理后, 得到文档中蕴含的空间实体集合, 称为空间参考集合 $R\{\{p_1, p_2, p_3, \dots, p_n\}, \{e_1, e_2, e_3, \dots, e_n\}\}$, 由水平空间参考子集和垂直空间参考子集组成. 这些集合中有核心实体, 也有辅助实体, 例如: “研究区头台油田位于黑龙江省大庆市肇源县境内, 在区域构造位置上处于松辽盆地中央拗陷区朝阳沟阶地西端, 为一向北西倾没的鼻状构造……”这段描述中“头台油田”核心实体, 而“黑龙江省”、“大庆市”、“肇源县”、“松辽盆地”、“中央拗陷区”、“朝阳沟阶地”等都是辅助性的描述实体, 是为了明确核心对象的空间背景而存在. SPIRIT^[31]、Geooreka^[32]等系统是将文档涉及的所有空间实体进行索引, 而张毅等^[9, 28]只对最重要的空间实体进行索引. 由于本领域的业务活动往往都有明确的地球空间主题约束, 产生的文档几乎都与明确的地球空间关联, 本研究认为对核心空间实体进行索引较符合领域需要.

空间聚焦本质是对空间参考集合 R 中元素的排序问题, 排序方法有基于地理参考树重要性度量和基于有向图的重要性度量, 本研究采用基于图排序的重要性度量算法, 其思想同网页排序 PageRank 算法^[33]类似, 节点的重要性取决于指向它的节点数量和它指向的节点数量, 指向它的节点数量越多且它指向的节点数量越少, 重要性得分就越高. 实体排序使用 Silva 等^[28]的图排序模型和收敛规则.

2.3 主题与空间多维混合索引结构

一个经典的查询包括 AOI、SOI 和兴趣主题 (theme of interest, TOI), 例如“查询杏树岗油田葡萄花油层关于非均质性的相关文档”. AOI 可以是一个地名, 例如“杏树岗油田”, 也可以直接是一个空间范围, 地名最终通过空间知识库转换为空间范围, 查询与经典的空间查询无异; 查询的 SOI 是一个地层名称, 例如“葡萄花”油层, 这种查询属于语义查询, 除了对查询有准确的语义解释外, 还有广泛的查询扩展需求, 如返回所有与“葡萄花”油层研究有关的文档, 以“葡 I 组”、“葡 II 组”油层组为空间主题的文档也在查询结果范围内; 业务主题是勘探开发业务人员关心的任何活动, 例如“非

均质性研究”,属于文本查询。

针对组合查询需求,本研究设计的索引结构采用的是先混合再组合模式,首先以 R^* -tree 为主索引倒排文件为辅索引建立平面空间索引 IR^* -tree,然后以地层本体为主索引,以倒排文件为辅索引建立垂向空间索引 $IOnto$ -tree, IR^* -tree 与 $IOnto$ -tree 是独立组合的,但是共用倒排文件结构,称为 $I-R$ - $Onto$ -tree. IR^* -tree 本质上还是一颗 R^* -tree,与倒排文件混合后实现“AOI+业务主题”查询,这种融合后的结构更紧致,包括 R^* -tree 的目录节点也要与倒排文件关联. $IOnto$ -tree 本质上是一颗基于地层本体结构的非平衡搜索树,会影响查询性能,其内在的查询扩展功能,加上哈希表辅助索引结构能弥补性能不足,其近似结构基于行政区划本体的索引在地理信息检索系统中得到实证^[34],其查询性能与 R -tree 持平或微低,与倒排文件混合后实现“SOI+业务主题”语义查询, IR^* -tree 与 $IOnto$ -tree 的查询结果进行交集运算,即最终查询结果。

倒排文件由两部分组成:词汇表和记录表集合.词汇表 $\{t_1, t_2, \dots, t_n\}$ 存储文档集中所有的不同词汇,记录表集合由若干记录表组成,记录表的模式是 $\langle id_d, score_{d,t} \rangle$,其中 id_d 表示包含词汇 d 的文档标识, $score_{d,t}$ 表示文档 d 与词汇 t 的相似度评分,该评分可以由 OKapi BM25 概率模型或者语言模型计算得出.以 IR^* -tree 为例,其非叶子节点 R_5 ,它在倒排文件 F_{R5} 记录表中记录为 $\langle d_{R5}, score_{d_{R5},t} \rangle$,其中 d_{R5} 表示由它的孩子节点组成的虚拟文档 $\{d_{R1}, d_{R2}\}$, $score_{d_{R5},t} = \max(score_{d_{R1},t}, score_{d_{R2},t})$,而 $score_{d_{R1},t} = \max(score_{d1,t}, score_{d2,t})$.通过虚拟文档可以预知叶子节

点中文档是否与查询匹配,以及最大文本相关性,提高剪枝效率,减小了查询空间,可以优化布尔最近邻查询和空间范围查询效率. $IOnto$ -tree 采用本体结构根据空间语义关系来分配节点,构建非平衡树,在查询时哈希表辅助结构提供直接取址方式,对于扩展查询利用其自身固有的语义结构直接搜索子树。

3 原型系统实现与测评

原型系统 PetroScope 在前期工作石油勘探开发企业搜索引擎^[25]的基础上构建,采用文档空间主题计算方法对文档处理组件进行了改造,其中文档的 NLP 预处理功能使用哈尔滨工业大学的语言技术平台,文档空间主题计算涉及的空间实体识别、空间实体消歧和空间实体聚焦则对原有组件进行了重写,三维索引结构中的倒排索引重用了 Apache Solr 的原有索引, IR^* -tree 索引在 SQLite3.9 基础上与倒排文件关联实现,基于地层本体的垂向空间索引 $IOnto$ -tree 使用 Java 从底层开发。

针对本研究的内容,评测任务主要包括领域文档的空间主题计算准确度和混合索引结构 $I-R$ - $Onto$ -tree 的性能评测.准备了两个文档集来测试本研究成果,文档集 EPDATA1 以领域学术文献和勘探开发生产报告为主,主要保留文档中的地学概述部分,业务主题涉及石油勘探开发 6 个业务域 1 000 多类业务活动;文档集 EPDATA2 从 Web 石油站点上抓取大量真实的网页. EPDATA1、EPDATA2 的基本参数见表 1, EPDATA1 用于测试空间主题计算方法, EPDATA2 用于测试 $I-R$ - $Onto$ -tree 索引结构。

空间主题计算模型评测,使用 PetroScope 对

表 1 测试文档集概况
Table 1 Documents dataset overview

属性	空间实体总数	每文档不重复词总数	文档集不重复词总数	文档集词总数
EPDATA1	120 367	224	31 847	15 076 604
EPDATA2	2 541 088	1 503	3 124 485	1 066 405 384

EPDATA1 整体标注,分 10 次随机抽取 100 个标注文档,确定文档的子领域后,由专业人员人工标注文档的水平空间主题和垂向空间主题,人工难以确定主题情况视为无效样本,结果见表 2。

索引结构的测试基线 R -tree 索引、基于地层本体的索引和倒排文件完全独立组合,即 3 个索引分别接收 3 种查询后,计算 3 个结果集的交集.查询集合从文档集的空间实体随机选择 4 组 100

表 2 空间主题计算评测结果
Table 2 Evaluation results of spatial reference extracting %

准确度	水平空间主题	垂向空间主题
最高	86.18	89.55
最低	73.66	80.26
平均	80.35	84.07

个水平空间实体作为水平空间查询词,查找与其至少一篇文档中有共现关系的 4 组 100 个垂直空

间实体和4组100个业务词分别作为垂直空间查询词和业务主题查询词, 保证每个查询都有结果. 实验主机CPU配置为Intel(R)Core i5-3230 2.60 GHz, 内存为4 GB. 以KNN查询类型为例, 分别对Top5、Top10、Top20、Top50进行比较, 运行时速度比基线索引结构最少提高25%, 最多提高了47.4%.

4 结论

本研究提出的空间主题计算模型准确度良好, 混合索引结构I-R-Onto-tree性能具有明显的改进. 地球科学领域文献普遍蕴含空间参考, 存在水平和垂向空间多维查询的需求, 所提出的框架、模型和索引结构经过调整后也可用于地球科学其他领域的空间信息检索.

由于文档空间主题计算的过程依赖于石油勘探开发领域空间本体, 为保证本体的准确性和完整性, 该本体的构建包括大量的人工参与, 下一步将研究该本体的半自动或者自动演化方法.

参考文献

- [1] Unneland T, Hauser M. Real-time asset management: from vision to engagement-an operator's experience [C]//Richardso N. Proceedings of the SPE annual technical conference and exhibition. Texas: the Society of Petroleum Engineers, 2005: 1-10.
- [2] 陈付平, 赵春宇, 凌雨. 石油勘探开发公共地理信息服务系统[J]. 计算机系统应用, 2011, 20(10): 15-19.
- [3] Sanderson M, Kohler J. Analyzing geographic queries [C]//Ross P, Chris J. Proceedings of the SIGIR workshop on geographic information retrieval. Sheffield: Association for Computing Machinery, 2004: 8-10.
- [4] 李丽双, 黄德根, 陈春荣, 等. 基于支持向量机的中文文本中地名识别[J]. 大连理工大学学报, 2007, 47(3): 433-438.
- [5] 唐旭日, 陈小荷, 许超, 等. 基于篇章的中文地名识别研究[J]. 中文信息学报, 2010, 24(2): 24-32.
- [6] 董华军. 基于条件随机场的地理命名实体识别方法研究[D]. 南京: 南京师范大学地理科学学院, 2009.
- [7] 鞠久朋, 张伟伟, 宁建军, 等. CRF与规则相结合的地理空间命名实体识别[J]. 计算机工程, 2011, 37(7): 210-215.
- [8] 乐小虬, 杨崇俊, 于文洋. 基于空间语义角色的自然语言空间概念提取[J]. 武汉大学学报: 信息科学版, 2005, 30(12): 1100-1103.
- [9] 张毅, 王星光, 陈敏, 等. 基于语义的文本地理范围提取方法[J]. 高技术通讯, 2012, 22(2): 165-170.
- [10] Andogah G, Bouma G, Nerbonne J. Every document has a geographical scop[J]. Data & Knowledge Engineering, 2012, 81: 1-20.
- [11] Borges K V, Davis C J R, Laender A F, et al. Ontology-driven discovery of geospatial evidence in web pages[J]. Geoinformatica, 2011, 15(4): 609-631.
- [12] 杜萍, 刘勇. 基于本体的中文地名识别[J]. 西北师范大学学报: 自然科学版, 2011, 47(6): 87-93.
- [13] Buscaldi D, Rosso P. Map-based vs. knowledge-based toponym disambiguation[C]//Proceedings of the 2nd international workshop on geographic information retrieval. California: Association for Computing Machinery, 2008: 19-22.
- [14] Buscaldi D. Approaches to disambiguating toponyms[J]. Sigspatial Special, 2011, 3(2): 16-19.
- [15] 朱少楠, 张雪英, 李明, 等. 基于行政隶属关系树状图的地名消歧方法[J]. 地理与地理信息科学, 2013, 29(3): 39-42.
- [16] Volz R, Kleb J, Mueller W. Towards ontology-based disambiguation of geographical identifiers[C]//Carey W. Proceedings of the 16th international world wide web conference. Banff: The International World Wide Web Conference Committee, 2007: 1-7.
- [17] Overell S, R Ger S. Using co-occurrence models for placename disambiguation[J]. International Journal of Geographical Information Science, 2008, 22(3): 265-287.
- [18] Campelo C E, Baptista C D. Geographic scope modeling for web documents[C]//Proceedings of the 2nd international workshop on geographic information retrieval. California: Association for Computing Machinery, 2008: 11-18.
- [19] Vaid S, Jones C B, Joho H, et al. Spatio-textual indexing for geographical search on the web[M]//Advances in spatial and temporal databases. Berlin: Springer, 2005: 218-235.
- [20] Zhou Ying-hua, Xie Xing, Wang Chuang, et al. Hybrid index structures for location-based web search[C]//Othein H. proceedings of the the 14th ACM international conference on information and knowledge management. Bremen: Association for Computing Machinery, 2005: 155-162.
- [21] Chen Yen-yu, Suel T, Markowetz A. Efficient query processing in geographic web search engines[C]//Clement Y. Proceedings of the the 2006 ACM SIGMOD international conference on management of data. Chicago: Association for Computing Machinery, 2006: 277-288.
- [22] Hariharan R, Hore B, Li C, et al. Processing spatial-keyword (SK) queries in geographic information retrieval (GIR) systems[C]//Ken B. Proceedings of the 19th international conference on scientific and statistical database management. Banff: The Institute of Electrical and Elec-

- tronics Engineers, 2007: 16-20.
- [23] de Felipe I, Hristidis V, Rishe N. Keyword search on spatial databases[C]//Alex B, Malu C. Proceedings of the IEEE 24th international conference on data engineering. Cancun: The Institute of Electrical and Electronics Engineers, 2008: 656-665.
- [24] Cong G, Jensen C S, Wu D. Efficient retrieval of the top-k most relevant spatial web objects[J]. Proc VLDB Endow, 2009, 2(1): 337-348.
- [25] Wang Zhi-bao, Wang Lin-lin, Shi Gui-ying. EISI-: an extensible security enterprise search system[C]//Shaohun Z. Proceedings of the computer science and network technology (ICCSNT), 2012 2nd international conference. Changchun: The Institute of Electrical and Electronics Engineers, 2012: 896-900.
- [26] Witmer J, Kalita J. Extracting geospatial entities from wikipedia[C]//David A E, Eduard H, Max M, et al. Proceedings of the semantic computing, 2009 ICSC '09 IEEE international conference. Berkeley: The Institute of Electrical and Electronics Engineers, 2009: 450-457.
- [27] Gelernter J, Balaji S. An algorithm for local geoparsing of microtext[J]. Geoinformatica, 2013, 17(4): 635-667.
- [28] Silva M J, Martins B, Chaves M, et al. Adding geographic scopes to web resources[J]. Computers, Environment and Urban Systems, 2006, 30(4): 378-399.
- [29] Liao W, Veeramachaneni S. A simple semi-supervised algorithm for named entity recognition[M]//Proceedings of the NAACL HLT 2009 workshop on semi-supervised learning for natural language processing. Boulder: Association for Computational Linguistics, 2009: 58-65.
- [30] Buscaldi D, Rosso P. A conceptual density-based approach for the disambiguation of toponyms[J]. International Journal of Geographical Information Science, 2008, 22(3): 301-313.
- [31] Purves R S, Clough P, Jones C B, et al. The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet[J]. International Journal of Geographical Information Science, 2007, 21(7): 717-745.
- [32] Buscaldi D, Rosso P. Geooreka: enhancing web searches with geographical information[C]// Antonellis V, Castano S. Proceedings of the seventeenth Italian Symposium on advanced database systems. Camogli: National Association for Database, 2009: 205-212.
- [33] Page L, Brin S, Motwani R, et al. The pagerank citation ranking: bringing order to the web[EB/OL]. 2008-12-28. [2017-04-12]. <http://ilpubs.stanford.edu: 8090/422/>.
- [34] Brisaboa N R, Luaces M R, Places Á S, et al. Exploiting geographic references of documents in a geographical information retrieval system using an ontology-based index[J]. Geoinformatica, 2010, 14(3): 307-331.

(责任编辑: 王春燕)