



## 基于MapReduce的书目数据关联匹配研究

虞为<sup>1</sup>, 陈俊鹏<sup>2</sup>

1. 南京大学信息管理学院 南京 210093;  
 2. 南京财经大学信息工程学院 南京 210023

Yu Wei<sup>1</sup>, Chen Junpeng<sup>2</sup>

1. School of Information Management, Nanjing University, Nanjing 210093, China;  
 2. School of Information Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China

- 摘要
- 参考文献
- 相关文章

Download: [PDF \(1137KB\)](#) | [HTML \(1KB\)](#) | Export: [BibTeX](#) or [EndNote \(RIS\)](#) | [Supporting Info](#)

**摘要** 提出一个基于MapReduce的书目数据关联匹配架构,通过参引MODS本体将MARC格式的书目数据转换成关联数据格式。再通过书目数据和书目数据间的关联匹配,以及书目数据和开放关联社区其他的关联数据间的匹配初步实现书目数据和其他关联数据集间的语义关联,使关联的书目数据成为关联开放数据社区中的一部分,为图书馆的知识发现和语义检索服务提供有效的语义数据支持。

**关键词:** [MapReduce](#) [关联匹配](#) [书目数据](#) [关联数据](#)

**Abstract:** In this paper, the MARC data is transformed to linked data, based on MapReduce model and MODS Ontology. Through the mapping among different linked open data sets, the library catalogue data can become part of the linked open data community and provide efficient semantic data to knowledge discovery and semantic service.

**Keywords:** [MapReduce](#), [Mapping and linkage](#), [Catalogue data](#), [Linked data](#)

收稿日期: 2013-06-06;

基金资助: 本文系国家自然科学基金项目“面向知识服务的知识组织模式与应用研究”(项目编号: 71273126)和国家自然科学基金项目“基于关联数据的图书馆语义云服务研究”(项目编号: 12CTQ009)的研究成果之一。

通讯作者 虞为 Email: yuw.nju@gmail.com

引用本文:

虞为, 陈俊鹏. 基于MapReduce的书目数据关联匹配研究[J]. 现代图书情报技术, 2013, V29(9): 15-22

Yu Wei, Chen Junpeng. Linking and Mapping of Library Catalogue Data Based on MapReduce[J], 2013, V29(9): 15-22

链接本文:

<http://www.infotech.ac.cn/CN/> 或 <http://www.infotech.ac.cn/CN/Y2013/V29/I9/15>


- [1] Heath T, Bizer C. Linked Data: Evolving the Web into a Global Data Space [M]. The 1st Edition. Morgan & Claypool Publishers, 2011.
- [2] Bizer C, Heath T, Idehen K, et al. Linked Data on the Web [C]. In: *Proceedings of WWW2008*, Beijing, China. 2008: 1265-1266.
- [3] The GeoNames Geographical Database[EB/OL]. [2013-07-12]. <http://www.geonames.org/>.
- [4] Samwald M, Jentzsch A, Bouton C, et al. Linked Open Drug Data for Pharmaceutical Research and Development [J]. *Journal of Cheminformatics*, 2011, 3(1): 19.
- [5] Bizer C, Lehmann J, Kobilarov G, et al. DBpedia - A Crystallization Point for the Web of Data [J]. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2009, 7 (3): 154-165.
- [6] Linking Open Government Data [EB/OL]. [2013-07-12]. <http://logd.tw.rpi.edu/>.
- [7] Malmsten M. Making a Library Catalogue Part of Semantic Web [C]. In: *Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications*. 2008: 146-152.
- [8] Summers E, Isaac A, Redding C, et al. LCSH, SKOS and Linked Data [C]. In: *Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications*. 2008: 25-33.
- [9] OCLC- WorldCat [EB/OL]. [2013-05-20]. <http://www.worldcat.org/>.

## Service

- ▶ 把本文推荐给朋友
- ▶ 加入我的书架
- ▶ 加入引用管理器
- ▶ Email Alert
- ▶ RSS

## 作者相关文章

- ▶ 虞为
- ▶ 陈俊鹏

- [10] 黄华军,曾新红,林伟明. OTCSS关联数据服务的研究与实现[J]. 现代图书情报技术,2012 (7/8): 40- 47.(Huang Huajun, Lin Weiming. Research and Implementation about Linked Data Service of OTCSS[J]. *New Technology of Library and Information Service*, 2012 (7/8):40-47.)
- [11] 欧石燕.面向关联数据的语义数字图书馆资源描述与组织框架设计与实现[J]. 中国图书馆学报,2012, 38(6):58-71. (Ou Shiyan. Design and Implementation of a Linked Data-oriented Framework for Resource Description and Organization in Semantic Digital Libraries [J]. *Journal of Library Science in China*, 2012, 38 (6):58-71.)
- [12] 夏翠娟,刘炜,赵亮,等. 关联数据发布技术及其实现——以Drupal为例[J]. 中国图书馆学报,2012, 38(1):49-57. (Xia Cuijuan,Liu Wei,Zhao Liang, et al. The Current Technologies and Tools for Linked Data: A Case of Drupal [J]. *Journal of Library Science in China*, 2012, 38 (1):49-57.)
- [13] 白海燕,朱礼军.关联数据的自动关联构建研究[J]. 现代图书情报技术,2010(2): 44-49. (Bai Haiyan, Zhu Lijun. Research on Automatic Interlinking of Linked Data[J]. *New Technology of Library and Information Service*, 2010(2): 44-49.)
- [14] DBLP [EB/OL]. [2013-05-20]. <http://dblp.uni-trier.de/>.
- [15] Moller K, Heath T, Handschuh S,et al. Recipes for Semantic Web Dog Food: The ESWC and ISWC Metadata Projects [C]. In: *Proceedings of the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference*. Berlin, Heidelberg: Springer-Verlag,2007: 802-815.
- [16] The Linking Open Data Cloud Diagram[EB/OL]. [2013-05-20]. <http://lod-cloud.net/>.
- [17] DuraCloud [EB/OL]. [2013-05-20]. <http://www.duracloud.org/>.
- [18] Dean J, Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters [C]. In: *Proceedings of the 6th Symposium on Operating Systems Design and Implementation*, 2004.
- [19] Bizer C. The Emerging Web of Linked Data [J]. *IEEE Intelligent Systems*, 2009, 24(5):87-92.
- [20] Oren E, Delbru R, Catasta M, et al. Sindice.com: A Document-oriented Lookup Index for Open Linked Data [J]. *International Journal of Metadata, Semantics and Ontologies*, 2008, 3(1): 37-52. 
- [21] Cheng G, Ge W, Qu Y. Falcons: Searching and Browsing Entities on the Semantic Web [C]. In: *Proceedings of the 17th International Conference on World Wide Web (WWW' 08)*.New York: ACM, 2008: 1101-1102.
- [22] MarOnt Ontology [EB/OL]. [2013-05-20].<http://deri.semanticweb.org/content/marcont-ontology>.
- [23] The FRBR Blog. Bibliographic Ontology Specification 1.0[EB/OL]. (2008-06-06). [2013-05-20]. <http://www.frbr.org/2008/06/06/bibliographic-ontology-specification-10>.
- [24] 白海燕,乔晓东. 基于本体和关联数据的书目组织语义化研究[J]. 现代图书情报技术,2010(9): 18-27. (Bai Haiyan, Qiao Xiaodong. Study of Semantic Bibliography base on Ontology and Linked Data[J]. *New Technology of Library and Information Service*, 2010 (9): 18-27.)
- [25] 王军,程煜华. 基于传统知识组织资源的本体自动构建[J]. 情报学报, 2009, 28(5): 651-657. (Wang Jun, Cheng Yuhua. An Automatic Approach to Ontology Building by Integrating Traditional Knowledge Organization Resources[J]. *Journal of the China Society for Scientific and Technical Information*, 2009, 28(5): 651-657.)
- [26] MOD Schema [EB/OL]. [2013-05-20]. <http://www.loc.gov/standards/mods/>.
- [27] SIMILE Widgets [EB/OL]. [2013-05-20]. <http://simile-widgets.org>.
- [28] RDF Ontology for MODS V3.1[EB/OL]. [2013-05-20]. <http://simile.mit.edu/2006/01/ontologies/mods3>.
- [29] Project Gutenberg Australia [EB/OL]. [2013-05-20]. <http://gutenberg.net.au/>.
- [30] code4rda [EB/OL]. [2013-05-20]. <http://code.google.com/p/code4rda/>.
- [31] The Library of Congress. Resource Description and Access [EB/OL]. [2013-05-20]. <http://www.loc.gov/aba/rda/>.
- [32] Bizer C, Heath T, Berners-Lee T. Linked Data - The Story So Far [J]. *International Journal on Semantic Web and Information Systems*, 2009, 5(3):1-22.
- [33] Volz J, Bizer C, Gaedke M, et al. Discovering and Maintaining Links on the Web of Data [C]. In: *Proceedings of the 8th International Semantic Web Conference*.Berlin, Heidelberg: Springer-Verlag,2009: 650-665.
- [34] DBpedia. DBpedia 3.8 Downloads[EB/OL]. [2013-05-20]. <http://wiki.dbpedia.org/Downloads38>.

- [1] 王忠义,夏立新,石义金,郑森茂.数字图书馆中层关联数据的创建与发布[J]. 现代图书情报技术, 2013,(5): 28-33
- [2] 刘炜,夏翠娟,张春景.大数据与关联数据:正在到来的数据技术革命[J]. 现代图书情报技术, 2013,(4): 2-9
- [3] 夏翠娟.RDB2RDF标准及应用研究[J]. 现代图书情报技术, 2013,(4): 10-17
- [4] 朱雯晶,夏翠娟,刘炜.SILK关联发现框架综析[J]. 现代图书情报技术, 2013,(4): 18-24
- [5] 钟远薪,李田章,刘炜.OPAC混搭关联数据应用研究[J]. 现代图书情报技术, 2013,(4): 25-29
- [6] 高劲梁,梁艳琪,马倩倩,周习曼,付旭雄.面向关联数据的引文知识链接模式研究[J]. 现代图书情报技术, 2013,29(3): 21-26
- [7] 洪娜,钱庆,范炜,方安,王军辉.关联数据中关系发现的可视化实践[J]. 现代图书情报技术, 2013,29(2): 11-17
- [8] 黄华军,曾新红,林伟明.OTCSS关联数据服务的研究与实现[J]. 现代图书情报技术, 2012,28(7/8): 40-47
- [9] 牛亚真,祝忠明.个性化服务中跨系统用户建模方法研究综述[J]. 现代图书情报技术, 2012,28(5): 1-6
- [10] 亢丽芸,王效岳,白如江.MapReduce原理及其主要实现平台分析[J]. 现代图书情报技术, 2012,28(2): 60-67
- [11] 黄永文,钱力.面向关联数据的信息检索服务研究综述[J]. 现代图书情报技术, 2012,(12): 2-8

- [12] 于倩倩, 李春旺. 注释服务研究述评[J]. 现代图书情报技术, 2012,(12): 9-14
- [13] 刘媛媛, 李春旺. 关联模型支持下的关联参考服务研究[J]. 现代图书情报技术, 2012,(12): 15-20
- [14] 牛亚真, 祝忠明. 个性化服务中关联数据驱动的用户语义建模框架[J]. 现代图书情报技术, 2012,(10): 1-7
- [15] 黄永文, 岳笑, 刘建华. 关联数据应用的体系框架及构建关联数据应用的建议[J]. 现代图书情报技术, 2011,27(9): 7-13