

DBRef: 研究心得的汇集与共享

贾琳琳 (Web 组)

1 什么是 DBRef

DBRef 是一个创新性的学术知识的发现、汇集与共享系统。它的目的是汇集整理研究领域的相关信息，尤其将能够体现研究者们研究经验、心得体会等信息集中总结，并提升为知识提供给所有研究者共享。它一方面能够自动地集成散布在 Web 环境中的相关信息；另一方面也鼓励研究者们积极参与，推荐自己在某个研究领域下的参考资料，重要信息，研究心得。在 DBRef 上用户可以自由发表个人学术观点，讨论热门话题，提出与解决问题……通过知识的汇集、挖掘与共享，DBRef 将架起一座桥梁，连接专家级研究者与入门级研究者，并最终为所有学术爱好者提供帮助。

2 为什么需要 DBRef

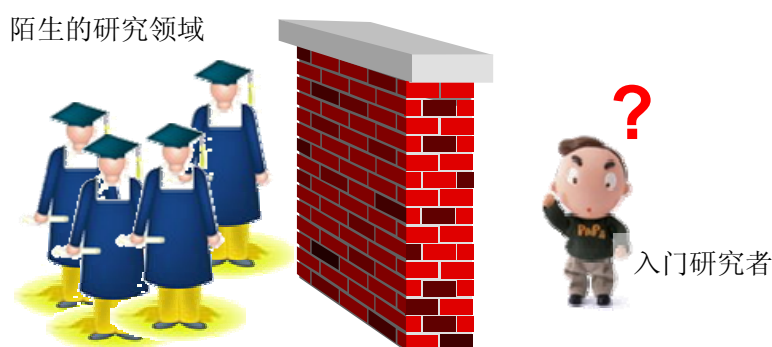


图 1 问题

一个有规律的现象

每一年进入实验室的研究生新生似乎都会问高年级博士们同样一些问题：能给我推荐一些这个领域的论文看吗？这个方向都有哪些著名会议？都有哪些人研究它呢？目前都有哪些方法？……似乎这些问题的答案是每一位入门级研究者的都急需的，不管在任何研究领域。类似的，每一年导师也要重复地给新生们推荐一些书籍，推荐一些课程，告诉他们如何看论文，如何写论文，如何学习，如何研究……

一个苦恼的问题

一个优秀的博士毕业了，如果能把他研究期间的经验与知识全部传授给其他低年级的同学该多好呢！

这些情况似乎在每一个研究组都存在。事实上，我们正是缺少一个有效的保存、整理、总结归纳以及查找这些通过时间与实践累积下来的经验知识的方法与工具，而这些知识零星散布在各处，关联性差，获取难，且极易丢失；而搜寻、聚集这些共同的资料可能会让一个研究者花费很多精力和时间，从而挤压了真正研究问题的过程。如果你是一位研究生，面对一个陌生的研究领域，最需要的是尽快了解这个研究领域的基础理论与研究概况，尤其是相关经典论文。你可以通过 Google 等搜索引擎查找相关信息，或者询问有经验的同学老师。显然，仅通过模糊的关键词进行查找，Google 的结果是庞杂而令人费解的；而他人的介绍也未必是详尽而具体的。即便是求助于 DBLP，面对 title 或者 author 搜索栏又该如何进行查询呢？这似乎不是件简单的工作，而且目前还没有任何很好的办法来解决它。

换另一个角度，如果你是该研究领域的一名较资深研究者，就会遇到不一样的问题。经过多年的研究与积

累，你对这个研究领域有了自己的一些学术观点，同时收集了一批有价值的论文。这些知识积累代表了你对这个领域的认识和理解，对于初入门者来说是极有帮助的，正是他们所急需的。同时，你也渴望将自己的学术观点同其他的研究者们分享讨论，并了解他们对什么感兴趣。

正是为了解决这些问题，我们尝试开发一个原型系统——DBRef，即“Database Reference”。现在，它正不断地更新与改进！你可通过此网址进行访问：<http://dbref.ruc.edu.cn/>。

3 DBRef 与相关的网站的区别

总结目前大家已熟知的相关网站，可以发现它们大致分为以下几类。以完整的论文元数据为主体的 DBLP；以论文引用为重点的 Citeseer；侧重于消息发布的 DBWORLD；提供学术动态信息的 DBLife。这些数据源从不同角度为用户提供一定的数据服务。然而，区别于它们，DBRef 强调的是如何汇集与挖掘学术领域的共享知识，而非单纯某一方面的信息检索与展现功能；强调关联性、重要性而非全面性。如果用户只是想要查找一个作者的论文列表和个人信息，他大可以求助于 DBLP 和作者主页，我们并不需要重复这些工作。我们强调的知识，应该定义为是用于解决问题的经过整理的易于理解和结构化的信息。DBLP 等网站能够为用户提供了一定信息，这些信息体现了它内部数据与数据之间的联系；而 DBRef 将更深一步汇集并整理这些信息，将信息与信息之间的联系挖掘出来总结为知识提供给用户。如果把数据、信息、知识之间的关系比喻为几何学上线、面、立体之间的关系，那么 DBRef 无疑将展现给用户一个更为立体的学术视图。

4 DBRef 框架与功能

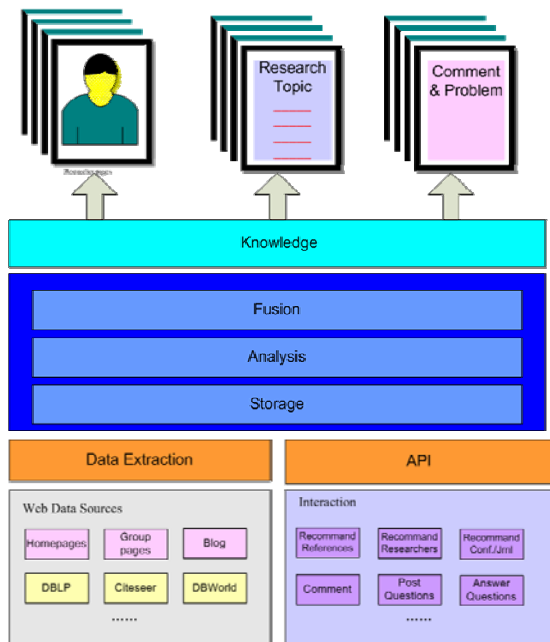


图 2 DBRef 框架图

DBRef 的数据来源分为两部分。一部分来自 Web 上的各种数据源，例如研究者的个人主页、博客、研究小组主页、会议主页以及 DBLP 等，对于不同类型的数据，DBRef 分别采用自动、半自动、手动的方法进行数据抽取；另一部分数据来自用户的参与、贡献，用户可以推荐相关资料、评论、提问与解答，DBRef 提供相应接口。对于汇集的各种数据，DBRef 进行有效组织与索引，通过关联性分析，将不同数据源中的信息重

新组织、融合、归纳，挖掘出有价值的信息提供给用户。例如用户可以根据其他人的推荐，发现一个领域的公认专家，最适合的、最权威的会议，相关领域的介绍和 Survey Papers，有价值的论文列表以及对于同一问题不同方法与技术的关系。

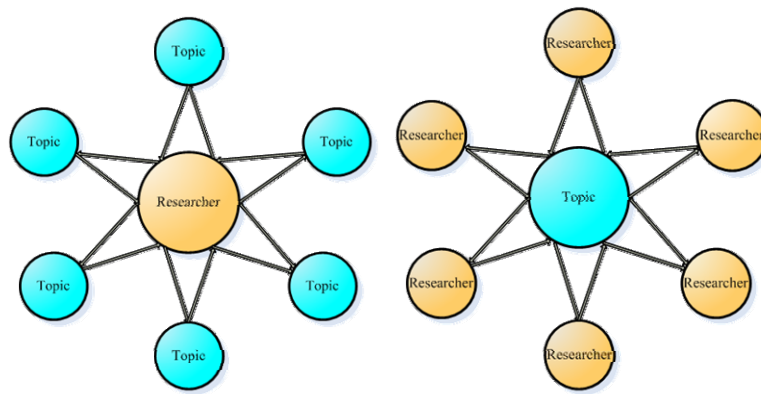


图 3 数据组织

在研究领域存在两个关键的对象：研究者和研究主题。研究领域一切相关的元素都是围绕这两者展开的。在它们之间存在 M: N 关系。一个研究者通常涉及多个研究主题，而一个研究主题可能有多个研究者研究。如何有效地展现这种多对多关系，并让用户方便地查找到相关联的信息是一个关键问题。在 DBRef 中，一方面以研究者为中心，为每位研究者或者注册用户建立主页面，主页面上会汇集他的个人信息，研究主题，主题介绍，不同主题下的推荐论文列表，相关会议与相关专家，个人研究心得等信息。另一方面以研究主题为中心，汇集这一主题下的研究者，相关论文，总体研究概括和各种方法的发展脉络和关系等信息。那么用户就可以从不同角度了解他所感兴趣的知识。

DBRef 从人、研究领域两个方面组织数据及相应功能，并突出焦点问题与热门讨论。给用户的主要功能有：

查找：相关研究主题与重要参考文献，survey，权威专家，著名会议，热门领域，最新主题，主题聚类、分层与变迁，不同研究主题下的相关概念、方法，交叉领域知识等。在 DBRef 上，你找到的论文不一定是引用率最高的（引用率具有一定时间累加性），也不一定是最知名的研究者的（研究界新星），甚至不一定是你的研究领域下的（交叉领域），但它恰恰是你最需要的。

发布与管理：作为 DBRef 的用户，研究者可以建立个人的研究主题分类结构，并上传他们收集的论文，同时发表自己的评论和学术观点。这些资源可以被 DBRef 的所有用户共享，这将会给新人们带来很大的帮助。对于同一个研究主题，不同的用户可以分别建立自己的论文列表，这些用户论文列表之间相互竞争，获得更多支持的便会排在更前面的位置，因为它被用户公认为是更有价值的论文列表。并且对于那些用户上传到本地的论文，DBRef 能够自动抽取论文元信息，免去了手动输入的麻烦，目前，自动抽取已达到相当高的准确率。

讨论与评论：可以对论文、观点、问题进行讨论与评论，交流观点与看法。可以发布问题也可以协助解决他人问题。虽然研究者们可以在会议上进行交流，但 DBRef 基于人人参与讨论的思想能够提供一个更自由的交流平台。

推荐与提醒：通过对用户个人信息以及的分析 DBRef 会为不同用户推荐个性化的信息与提醒功能。

上述功能在目前的 DBRef 中并没有完全实现，在今后各个功能会逐一上线。也欢迎大家建议新的功能。

5 结束语

DBRef 其实是一项公益性的工作，它的诞生虽源于 WAMDM 实验室每一次人员流动过程中出现的小小窘境，但它的将来绝对不仅仅只属于一个实验室或者一些人，它所沉淀下来的宝贵的知识财富将能够惠及所有研究者和任何渴望知识的人，希望它能够让不同国家、不同地区、不同组织的但拥有共同研究爱好、共同需求的人们走到一起，无障碍地进行交流与分享，让智慧的火花在这里迸发，让知识与思想跨越鸿沟，带动所有人前进。