

# 基于大数据分析的技术转移平台<sup>①</sup>



张颖<sup>1</sup>, 管震<sup>2</sup>, 蒋浩<sup>3</sup>

<sup>1</sup>(江苏省生产力促进中心, 南京 210042)

<sup>2</sup>(太仓中科信息技术研究院, 太仓 215400)

<sup>3</sup>(中国科学院计算技术研究所, 北京 100190)

通讯作者: 管震, E-mail: tcguanzhen@163.com

**摘要:** 随着国家对技术转移工作的大力支持, 越来越多的技术转移平台开始涌现. 但是当前的大部分技术转移平台对供需数据无法做到很好的匹配, 所以平台对于技术转化成功率的提升效果受到一定限制. 为了解决这个问题, 本文提出一种新颖的技术转移平台实现方法, 它主要利用大数据挖掘技术来整理和分析平台数据, 在此基础上, 我们结合全文检索、数据采集、RESTful 接口等其他技术, 共同提升数据的匹配度, 扩大数据的展示范围, 从而提升技术转化率. 目前平台已在江苏省部署运行, 通过平台促成了大批技术成果转移落地.

**关键词:** 大数据; 技术转移; 数据采集; 全文检索; RESTful 接口

引用格式: 张颖, 管震, 蒋浩. 基于大数据分析的技术转移平台. 计算机系统应用, 2020, 29(7): 110-116. <http://www.c-s-a.org.cn/1003-3254/7477.html>

## Platform of Technology Transfer Based on Big Data Analysis

ZHANG Ying<sup>1</sup>, GUAN Zhen<sup>2</sup>, JIANG Hao<sup>3</sup>

<sup>1</sup>(Productivity Centre of Jiangsu Province, Nanjing 210042, China)

<sup>2</sup>(Taicang-CAS Institute of Information and Technology, Taicang 215400, China)

<sup>3</sup>(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** With the strong support of China for technology transfer, more and more technology transfer platforms have been emerging. However, it is difficult to match supply and demand on existing platforms. Therefore, the platform's potential in improving the success rate of technology transfer is limited. For solving this issue, this study proposes a novel implementation method for technology transfer platform which uses big data mining technology to organize and analyze platform data. We combine other technologies such as full-text search, data collection, and RESTful interface to jointly improve the matching degree of the data, expand the visible range of the data, and ultimately improve the technology transfer rate. At present, a platform has been deployed and operated in Jiangsu Province. Through the platform, a large number of technology transfers have been implemented.

**Key words:** big data; technology transfer; data acquisition; full-text index; RESTful interface

## 1 引言

技术转移平台是一种用于辅助科技成果转移落地的公共服务平台, 平台以技术集成为主体, 同时集成信

息、资金、中介服务等多种技术成果转移相关信息. 技术转移平台实现了资源、资讯的聚集与服务的共享, 同时通过企业和市场的需求来指导技术的发展和方向.

① 基金项目: 江苏省产学研合作移动智能对接信息平台及服务能力建设 (BM2015015)

Foundation item: Mobile Intelligent Docking Information Platform and Service Capacity Building for Industry-University-Research Cooperation of Jiangsu Province (BM2015015)

收稿时间: 2019-12-03; 修改时间: 2019-12-27; 采用时间: 2020-01-02; csa 在线出版时间: 2020-07-03

“十三五”规划以来,我国一直在深入贯彻落实《促进科技成果转化法》等一系列的相关法律法规,我国经济的发展以及科技与经济的结合都离不开科技成果的转移转化<sup>[1]</sup>。

目前,我国的技术转移平台主要分为民营性质平台和国家、地方支持的公益平台。截止2016年底,我国30个省(自治区、直辖市(除西藏))、新疆生产建设兵团和5个计划单列市已经开始布局建立技术转移示范机构,其中江苏是拥有创新资源最丰富、技术转移最活跃的省市之一,已建成45家技术转移平台<sup>[2]</sup>。虽然拥有了大量的技术转移平台,但是平台上的供需匹配却并不高,文献<sup>[3]</sup>中指出目前所有平台上的供需数据只有约20%可以做到比较好的匹配,如何提升平台供需数据的匹配率成为了各个技术转移平台的进一步建设方向。对此有许多新的技术转移转换的平台建设方案被提出,如采用云服务环境建立技术转移平台<sup>[4]</sup>,但是目前技术转化率不高的问题依然存在。

近年来国内外对基于大数据分析的技术平台进行了一些研究,基于大数据分析技术的平台在各个领域得到的成功的应用,如文献<sup>[5]</sup>中将大数据分析技术应用于气瓶生产的安全进行管理。基于大数据分析技术的平台其特点在于对数据的获取、清洗、集成、分析和可视化等多个模块通过松耦合的方法进行组装<sup>[6,7]</sup>,并实现传统关系型数据库无法解决的海量数据在线查询的问题<sup>[8,9]</sup>,从而让平台与业务更加的契合。因此,通过大数据分析技术设计并建立一个针对技术转移转换相关数据的数据管理平台,可以有效地解决当前技术转移转率不高的问题。

## 2 需求与现状

目前,大部分技术转移平台主要提供简单的数据存储以及查询展示,并没有将平台上的海量数据的价值充分的发挥,科技成果的提供方与需求方之间的信息障碍仍然存在,并且严重阻碍了技术的转移。以2016年为例,江苏省全省高校专利授权19 848件,专利技术转让只有687件,仅占专利授权的3%<sup>[10]</sup>。通过对多个技术转移平台的分析发现存在以下几个问题:

1) 大部分的技术转移平台的检索条件都比较简单,只是实现了常见的模糊查询,并没有对查询条件以及各类资源信息准确的语法和语义上的分析,导致查询结果不太容易准确匹配用户需要。

2) 技术转移平台中存储的海量数据并没有被充分

利用起来,数据中存在的潜在价值需要被深入挖掘,例如用户的浏览记录、咨询记录等数据可以从中挖掘出用户关心的资源信息,从而提高用户查询数据的效率,降低供需双方的信息壁垒。

3) 没有考虑到地理环境因素对技术转移的影响。在实际的环境中技术的对接在很大程度上受到区域政治、经济和地理位置等地理环境因素的影响<sup>[11]</sup>。但是当前大部分的平台中在资源匹配时并没有考虑到这个因素。

此外当前各级区县、地级市建立了大量的技术交易平台,但相互之间的数据并不互通,导致数据分散、冗余,管理分析困难。为此江苏省政府2018年提出要加强科技成果交汇<sup>[12]</sup>的政策,希望通过对科技成果数据的统一汇总、分析,实现科技成果的转移效率的提升,加快地方科技与经济的发展。

总结上述问题,新的技术转移平台系统需要满足以下几点需求:

1) 搭建一个技术转移相关资源的发布、展示、查询和对接的平台,能够让专家、企业、高校在平台上发布各类与技术转移相关的资源信息,包括技术成果、技术需求、专家团队、技术专家等。对于这些资源系统能够提供条件丰富的检索功能,从而满足用户各个方面的查询需求,此外能够提供如智能提示等功能,对用户的查询进行一定的帮助和指引。

2) 对系统中的资源数据实现数据深度挖掘,通过数据挖掘分析用户行为和资源信息,分析用户的相似性和资源信息的相似性,实现对用户的定制化信息推荐,帮助用户快速定位其需要的技术转移资源,提高技术转化的成功率。

3) 对于所有的资源信息绑定其地区与位置信息,给用户地区查询功能,同时结合手机APP的实时定位功能,让用户能够方便的查询其附近的资源信息,解决地理环境因素对技术转移的影响。

4) 建立一个数据中心,通过Web API的方式实现数据互通。各级区县、地级市的平台数据可以通过Web API将其平台上的所有资源信息上传至数据库中心。同时各级区县、地级市的平台数据也可以通过Web API获取数据库中心的数据挖掘结果以及数据查询结果,从而降低各级区县、地级市搭建平台的复杂度。最终让数据能够统一汇总、分层下发。

5) 数据采集与发布,使用数据采集工具从互联网上采集各类与技术转移相关的信息,如国家政策、新闻资讯等提供给各类用户进行查看。数据采集完成之

后能够对采集到的数据进行智能化的清理,将数据整理成符合平台数据库格式要求的数据并自动入库提供给用户查看。

### 3 系统总体设计

目前常见的系统体系结构有 C/S 结构和 B/S 结构。B/S 架构 (Browser/Server, 浏览器/服务器模式), 是 Web 兴起后的一种网络结构模式。采用 B/S 结构的系统其核心功能都集中在服务器上, 用户只需要通过浏览器即可与系统进行交互, 十分符合本系统用户分散的特点, 因此本系统采用 B/S 结构。

随着智能手机的普及, 用户越来越希望能够通过智能手机来访问系统, 因此本系统除了像传统 B/S 结构通过客户端的浏览器与系统进行交互的方法外也提供了通过 APP 与系统进行交互的方式。

此外系统还提供 Web API 给其他区县、地级市的技术转移平台与本系统的数据进行互通, 实现数据的统一汇总和分层下发。

数据采集服务器实现数据的采集并录入数据库服务器中。文件分布式服务器用于存储系统的所有文件, 如图片、附件、全文索引等, 从而实现逻辑业务与文件读取的分离, 提高平台的吞吐量和可扩展性。大数据分析服务器提供大数据分析服务。系统物理结构图如图 1 所示。

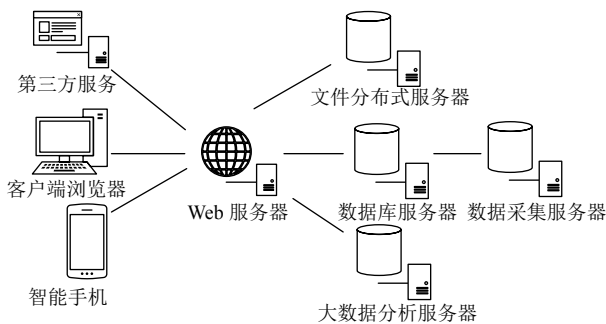


图 1 系统物理结构图

### 4 系统设计与关键技术

根据对技术转移转换平台业务分析的结果技术转移平台系统整体被划分为 4 个子系统, 分别是业务子系统、采集子系统、Web API 子系统和大数据分析子系统。在数据存储方面平台系统有 3 个数据存储模块, 分别是关系数据存储模块采用 MySQL 数据库, 全文检索引擎和大数据存储管理。软件结构示意图如图 2 所示。

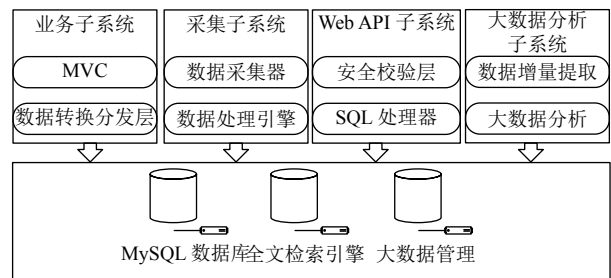


图 2 软件结构示意图

#### 4.1 业务子系统

业务子系统主体以微软的 .NET Framework 4.0 为平台, 采用 ASP.NET MVC 4 的框架, 将底层数据、业务逻辑与展示页面进行有效的分离, 从而降低系统的耦合性。此外, 本系统在 MVC 的三层架构的基础上采用面向接口的设计, 通过控制反转的方式实现系统中各个部件之间的解耦与动态调用。

数据访问驱动层采用开源的 ORM (Object Relational Mapping) 框架 petapoco。相比于其他的 ORM 框架, petapoco 的优势在于轻量级、高性能、低耦合、易扩展<sup>[13]</sup>。但是原生的 petapoco 并不支持对全文检索引擎和大数据存储的读写操作, 为此平台通过 AOP 的方法设计了一个数据转换分发层, 实现对 petapoco 的横向扩展, 从而实现对全文检索引擎和大数据存储的读写功能。

#### 4.2 采集子系统

数据采集器采用网络信息资源采集技术对互联网上的数据进行采集, 主要采集政策新闻类和科技成果类数据。

数据被采集后将先存入高速队列中, 然后由数据处理引擎对数据进行异步的处理然后入库。数据处理引擎对数据进行两个方面的处理: (1) 除去多余的数据如 html 标签; (2) 根据数据库结构提取每个字段的数据。

采集系统采集的数据不会直接进入业务子系统的数据库中, 而是存入采集子系统的临时库中, 然后由人工对采集数据进行最终的筛选并转入业务子系统的数据库中。

#### 4.3 Web API 子系统

Web API 子系统采用 RESTful 架构设计实现。RESTful 架构的优势在于其架构遵循 CRUD 原则, 可以通过统一资源标识符 (Universal Resource Identifier,

URI) 来识别和定位资源, 并且针对这些资源而执行的操作是通过 HTTP 规范定义的, 从而可以简化开发<sup>[14]</sup>. 而返回的状态和数据是通过 HTTP 响应码和 JSON 这样的统一规范来实现, 因此能够方便的给第三方系统进行调用.

在 API 的入口和数据库查询前分别设置了安全校验层和 SQL 处理器. 安全校验层是为了保证 Web API 请求的安全性, 对于每一个申请 Web API 的用户系统都会通过文献<sup>[15]</sup>中的方法生成一个 Token 用于验证访问的合法性, 此外通过解析 Token 的方式定位本次请求的用户身份, 然后通过 IOC (Inversion Of Control) 的方式自动地去调用 SQL 处理器, 通过对 SQL 的二次处理实现对该用户数据访问的权限控制.

#### 4.4 大数据分析子系统

大数据分析系统定时从大数据存储系统中提取数据并进行分析. 分析操作主要分析文本相似性和用户相似性两种数据.

文本相似性分析是分析当前数据库中所有文本数据的相似度并将计算所得的相似的数据写回大数据存储系统中, 后期查询相似数据时只要通过相似值进行排序查询即可.

用户相似行分析是通过用户浏览信息、对接信息等操作信息对用户的相似性进行分析, 将用户的相似性结果记入大数据存储系统中. 后期将相似的用户间浏览的数据推荐给用户.

#### 4.5 关键技术分析

大数据分析: 大数据分析采用 TF-IDF 算法<sup>[16]</sup>和协同过滤算法<sup>[17]</sup>来实现相似性推荐和用户行为分析. 具体步骤是首先通过 TF-IDF 算法从资源数据中提取出若干个关键字组成关键字向量, 然后使用协同过滤算法计算两个文献关键字向量的余弦值得出两个向量的相似度, 超过指定阈值的将被判断为相似. 而当用户浏览某个资源时, 系统会显示与其相似的其他资源. 用户喜好分析通过分析用户最近的浏览记录、对接记录和查询记录来计算用户的相似度, 将相似用户中的热点数据推以邮件、短信和 APP 通知的方式向用户推送.

Lucene 搜索引擎: Lucene 是一款高性能的、可扩展的信息检索工具库, 其内核通过倒排索引的方式实现高效的全文检索, 此外通过配置分词器和过滤器的方式可以方便地实现复杂的定制化功能, 如同义词检索、智能提词等<sup>[18]</sup>.

SPATIAL 索引: MySQL5.7 开始支持 SPATIAL 索引, 通过这种索引模式可以实现对地理空间位置的查询, 能够帮助手机端实现查看当前地理位置附件的信息的功能.

数据分区: MySQL 的数据分区是将数据按照给定的规则分配在不同的数据文件中, 这样当查询同一个分区的数据时可以提升查询效率. 平台通过定时任务定时地对对接记录、访问记录等数据量大的数据进行按时间的分区操作, 这样当用户查询最近的相关数据时能大大提升查询的效率.

移动端技术: 系统的移动端应用开发采用了 Google 的 MVP (Model-View-Presenter) 框架, 从而实现模型与视图完全分离, 解决传统移动端应用开发中 Activity 代码臃肿、耦合度高的问题<sup>[19]</sup>. 此外在移动端程序中还采用了 Glide 框架实现异步加载各类图片文件以及 RxJava 框架优化 HTTP 请求代码和步骤, 从而提升移动端程序的响应速度.

## 5 系统功能设计

根据需求设计出符合需求分析的系统, 系统的主要功能模块包括: 用户管理、对接管理、大数据分析、Web API 管理和系统管理五大模块, 系统整体功能结构如图 3 所示.

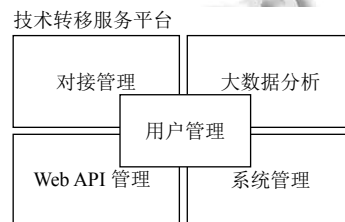


图3 系统整体功能结构示意图

用户管理模块除了登录、注册外主要实现了用户的角色申请和审核功能, 用户通过申请可以获得不同的角色, 每一个角色拥有不同的平台权限. 平台中主要有 7 类角色, 分别是专家、企业、高校、科技局、第三方服务机构和管理员. 其中专家、企业和第三方服务机构是平台的主要角色, 可以上传各类技术转移资源并与资源申请对接. 而高校、科技局和管理员是平台的管理类角色, 高校负责专家的审核管理, 科技局负责企业的审核管理, 而管理员负责对所有角色的用户进行审核管理.

对接管理模块实现了各类技术转移资源信息的发

布、查询和对接的功能,平台内将技术转移资源分为了科技成果、技术专家、专家团队、研发需求、技术服务五类.用户可以向需要的资源发起对接请求,资源所有人收到请求并同意后就完成了一次技术转移的对接.

基于大数据的技术转移服务平台对比其他相似的技术转移平台最大的区别有两个方面:第一,通过大数据分析的方法将存储在系统中的庞大数据信息充分的利用,通过系统分析加快对接目标的定位;第二,通过建立数据中心,采用 Web API 的方式汇总数据,解决数据冗余、分散的问题,并为大数据分析系统提供足够的数据库.

Web API 实现数据的统一汇总和分权查询.汇总功能提供的插入和修改两个接口,第三方系统调用接口可以实现数据的插入和修改.分权查询功能提供了一个查询的接口,当第三方系统调用接口时,接口将根据第三方提供的查询条件返回查询结果.第三方调用 Web API 前需要提前注册,系统会为其分配校验的 TOKEN,同时建立其数据访问权限所对应的数据库视图.当调用查询接口时,系统通过分析 TOKEN 确定当前调用的第三方身份,然后调用其对应的视图进行查

询,从而实现数据的分权查询.

系统管理模块实现系统日志记录和消息推送功能.系统日志包括用户的操作日志和异常日志.消息推送功能包括邮件发送、短信发送和 APP 推送,并且为了提升消息推送的响应速率,系统采用消息队列的方式实现消息的异步推送.

## 6 系统实现及应用

本系统采用 C#语言开发,在 .NET 框架下完成系统功能开发,服务器采用 Windows Servers 2012+IIS8,数据库采用 MySQL 数据库,整体采用 B/S 架构,客户端向下兼容至 IE8 浏览器,方便用户跨平台使用.

本系统目前已部署在江苏省运行,为江苏全省提供技术转移对接服务,平台各参与方可以自主完成技术需求和科技成果的对接,同时还能通过多种方式参与各种产学研活动、及时了解产学研资讯信息,通过移动端 APP 用户还可以随时随地参与产学研对接各环节. Web 端的页面展示效果如图 4 所示,移动端 APP 的展示效果如图 5 所示,其中右图展示了基于地理位置的查询和地图展示效果.



图4 Web端功能界面

基于大数据分析系统,用户可以方便地进行智能检索,且能根据用户活动推荐其可能感兴趣的数据资源,通过多种方式帮助用户高效的实现技术需求和技

术成果的对接.如图 6 所示,在用户进行检索时,平台会智能提示可能需要的查询词,检索操作会同时对数据资源的正文和标题进行全文检索,无需人工标引即

可对信息源的完整内容进行匹配. 如图 7 所示, 平台可以根据用户正在浏览的资源, 提供与其相似的数据资源并推荐, 还能根据用户的查询和浏览记录分析其可能感兴趣的数据资源并加以推荐, 引导用户发现自己所需的数据资源.



图 5 移动端 APP 功能界面

自平台上线至今 3 年内已累计注册了用户 5 万多人, 其中专家有 1 万多人, 企业近 2 万家, 发布科技成

果、技术需求等信息约 6 万多条, 有效对接次数 8 千余次, 促使江苏省的技术转移对接率得到了大幅的提升.



图 6 全文检索与智能提词界面



图 7 相似推荐以及基于用户兴趣的推荐

### 7 总结

技术转移困难是一个长期存在的问题, 本系统依托于大数据分析, 通过大数据挖掘充分分析用户行为和资源数据, 挖掘出系统中的资源和用户的相似性信息, 从而为用户提供精确的信息推荐, 帮助用户快速定位其需要的资源信息, 提高技术成果的转化率. 同时,

通过全文检索、智能提词、地区位置检索等查询方式使用户能够更加方便地对技术转移资源信息进行查询. 此外, 通过 Web API 的方式实现平台数据与各级区县、地级市平台数据的上下互通, 从而让数据能够统一汇总、分层下发, 从而解决技术转移数据分散和冗余的问题. 平台自上线运行以来, 目前在平台上持续转

移落地技术成果并取得了一定的成效,从江苏省的实践可以看出,平台对于技术转移的促进以及对接效率的提升有着较大的帮助。

### 参考文献

- 1 林超辉,关凯胜,陈辉,等.搭建基于大数据分析的高校技术转移服务平台.中国高校科技,2017,(S2):110-112.
- 2 孙红军,李红,徐超.京藏技术转移服务平台建立的思考与建议.绿色科技,2018,(8):249-250.
- 3 杨德林,夏青青,马晨光.在线技术转移平台的供需匹配效率分析.管理科学,2017,30(6):104-112. [doi: 10.3969/j.issn.1672-0334.2017.06.008]
- 4 肖国华,唐衢,王江琦.云计算环境下专利技术转移平台研建设计.情报杂志,2014,33(10):153-158. [doi: 10.3969/j.issn.1002-1965.2014.10.026]
- 5 吕根源.基于大数据分析的气瓶安全监管平台设计与实现.中国计量,2017,(10):36-38.
- 6 王建民.领域大数据应用开发与运行平台技术研究.软件学报,2017,28(6):1516-1528. [doi: 10.13328/j.cnki.jos.005231]
- 7 杜小勇,卢卫,张峰.大数据管理系统的历史、现状与未来.软件学报,2019,30(1):127-141. [doi: 10.13328/j.cnki.jos.005644]
- 8 Song SX, Zhang AQ, Wang JM, *et al.* SCREEN: Stream data cleaning under speed constraints. Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. Melbourne, Australia. 2015. 827-841.
- 9 Wang JM, Song SX, Lin XM, *et al.* Cleaning structured event logs: A graph repair approach. Proceedings of 2015 IEEE 31st International Conference on Data Engineering. Seoul, Republic of Korea. 2015. 30-41.
- 10 成曦,徐欣,巢俊.加强建设江苏省技术转移体系的若干思考.江苏科技信息,2018,35(19):7-9. [doi: 10.3969/j.issn.1004-7530.2018.19.003]
- 11 陈红喜,彭鹏程,戴书春,等.江苏技术转移中高校科技成果转化问题的数据包络分析.中国科技信息,2015,(16):144-146. [doi: 10.3969/j.issn.1001-8972.2015.16.046]
- 12 江苏加快推进全省技术转移体系建设.政策瞭望,2018,(7):54.
- 13 Brad Robinson>Welcome to the official PetaPoco repository. <https://github.com/CollaboratingPlatypus/PetaPoco>.
- 14 刘润杰,龙华秋,容振邦.基于RESTful架构的任务发放平台开发.网络安全技术与应用,2019,(9):41-43. [doi: 10.3969/j.issn.1009-6833.2019.09.023]
- 15 廖露阳,郭兵.基于Android App安全登录认证解决方案.现代计算机,2016,(35):9-12. [doi: 10.3969/j.issn.1007-1423.2016.35.002]
- 16 曹姗.基于TF-IDF特征提取的短文本分类方法.工业控制计算机,2018,31(4):109-110. [doi: 10.3969/j.issn.1001-182X.2018.04.045]
- 17 姚婷.基于协同过滤算法的个性化推荐研究[硕士学位论文].北京:北京理工大学,2015.
- 18 沙阳阳,吴陈.基于Web的Lucene全文搜索排序算法的研究.计算机与数字工程,2019,47(5):1208-1211,1239. [doi: 10.3969/j.issn.1672-9722.2019.05.038]
- 19 倪红军.基于MVP模式的Android应用开发研究.电子设计工程,2018,26(11):6-9,13. [doi: 10.3969/j.issn.1674-6236.2018.11.002]