

EPS 知识服务平台介绍

北京福卡斯特信息技术有限公司

2016 年 8 月

目录

一、产品简介.....	3
二、资源检索系统.....	3
1、云分析平台.....	4
1.1 数据可视化.....	4
1.2 分析预测.....	4
1.3 个人数据管理.....	6
2、智能报告生成平台.....	6
2.1 原始数据表格处理.....	6
2.2 表格分析.....	6
2.3 智能报告生成.....	7
三、主题检索系统.....	8
1、知识发现平台.....	8
2、文献计量分析.....	8
2.1 研究趋势.....	8
2.2 研究热点.....	9
2.3 热点期刊.....	9
2.4 热点学者.....	9
2.5 热点学科.....	9
2.6 热点机构.....	9
2.7 热点基金.....	9
3、知识图谱.....	10
3.1 组合衍生词图谱.....	10
3.2 机构合作图谱.....	10
3.3 机构+关键词图谱.....	10
3.4 共现关键词图谱.....	10
3.5 作者合作图谱（显性）.....	11
3.6 作者合作图谱（隐性）.....	11

一、产品简介

EPS 知识服务平台采用全新数据服务模式，满足用户对数据更深层次的需求，其宗旨在于为广大读者提供在实证研究过程中存在的“从查阅文献到收集数据、从利用数据到得出结论”环节中所缺失的数据发现与数据分析服务。

EPS 知识服务平台基于知识发现、应用集成、开放互联的设计理念，摒弃数据资源的简单堆砌，强调数据利用，通过云分析平台和智能报告生成平台，将一系列与数据密切关联的应用集成到一个平台中。通过对文献中指标的提取与挖掘分析，发现主题与指标间的关联关系，形成知识发现图谱。提供知识分享、数据上传等个性化服务，实现不同类型数据资源间的互通互联。



二、资源检索系统

EPS 知识服务平台基于面向海量数据的智能搜索引擎技术，进行全面化的最优匹配检索，可在毫秒时间内精确检索出千万级别的数据，不但实现了时间序列、统计表格、文献元数据三种类型资源的同时快速检索，还能实时的进行动态查询、多维度、多频度的数据聚合分类以及统计。资源检索对于时间序列提供频度、地区、领域、来源、出版物、标签等六个维度的分类聚合，对于表格资源提供频度、区域、时间、出版物等四个维度的分类聚合，对于文献资源提供年份、作者、学科、期刊、机构、基金等六个维度的分类聚合。

1、云分析平台

EPS 知识服务平台采用 SaaS（软件即服务）模式，通过云分析为用户提供高质量、高效率、低成本的数据处理、可视化展现、分析挖掘等软件服务以及结果的共享与储存。同时，用户可以在云端上传个人数据进行分析处理，满足用户个性化需求，为其科学研究或论文撰写提供专业强大的工具支持。

1.1 数据可视化

数据可视化多样的可视化展示途径，分别包括智能表格、统计图表、行政地图。

其中智能表格是统计图表和行政地图的基础，分为五大区块：指标操作、功能菜单、用户操作、可视化区、表单操作；此外，系统还提供电子表格的在线重现（上传功能）。

统计图表是 EPS 知识服务平台对数据的另一类可视化展现方式，共有折线图、面积图、柱形图、条形图、饼图、环形图、雷达图、散点图等 13 种图表。系统的图与表格具备联动展示特效，能够展示用户选中的任意数据区域。

地图是 EPS 知识服务平台对数据的高级可视化展现方式，有中国地图、世界地图两类。地图具备展示指标切换、背景色系选择、时间序列动态播放、数据钻取等功能，满足用户对数据地图的绝大多数需求。

1.2 分析预测

数据的价值在其背后隐藏的规律，EPS 知识服务平台为用户提供了功能强大的分析预测模型库。包括描述性统计、变量处理、缺省值处理、时间聚合、回归分析、相关分析、时间序列分析、主成分/因子分析、高级计算器等 9 大模块，共计 40 余类常见分析模型，涉及时间序列数据、截面数据及面板数据等多种数据模型的计算处理。

描述性统计分析是对一组数据的各种特征进行分析，以便于描述测量样本的各种特征及其所代表的总体的特征。描述性统计分析的项目很多，常用的如平均数、标准差、中位数、频数分布、正态或偏态程度等等。

变量处理模块主要用于对时间序列变量进行初步简单处理，主要包含了增长率、差分、对数转换、滞后函数、先行函数、发展速度、累积方法、标准化等。

缺省值处理模块用于对时间序列变量存在的缺省值进行相关处理，主要包含了删除、序列均值、三次样条内插、几何插补、线性插值、线性趋势、前一个值、后一个值、前 N 个增长率、后 N 个增长率、相邻 N 点均值、相邻 N 点中位数、随机值。

时间聚合的功能主要用于考察时间序列的周期特性，用于寻找某个周期（如年度、半年度、季度）内该序列的最大值、最小值、首值、末值、平均值、标注差、和值。该算法仅对时间序列有效，对月度与季度两种频度数据有效，可实现从高频度向低频度数据的转化。年度数据不能进行聚合处理。

回归分析是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。运用十分广泛，回归分析按照涉及的自变量的多少，分为回归和多重回归分析；按照自变量的多少，可分为一元回归分析和多元回归分析；按照自变量和因变量之间的关系类型，可分为线性回归分析和非线性回归分析。结合平台特性，EPS 知识服务平台当前提供三大类回归分析方法，分别是线性回归、曲线估计、二阶段最小二乘回归，其中线性回归及曲线估计又包含多种方法级多种模型。EPS 知识服务平台允许线性回归应用截面数据模型与时间序列数据模型。相关分析是研究现象间是否存在某种依存关系，并对具体有依存关系的现象探讨其相关方向以及相关程度，是研究随机变量之间的相关关系的一种统计方法。

时间序列是按时间顺序的一组数字序列。时间序列分析就是利用这组数列，应用数理统计方法加以处理，以预测未来事物的发展。时间序列分析是定量预测方法之一，它的基本原理：一是承认事物发展的延续性。应用过去数据，就能推测事物的发展趋势。二是考虑到事物发展的随机性。任何事物发展都可能受偶然因素影响，为此要利用统计分析中加权平均法对历史数据进行处理。该方法简单易行，便于掌握，但准确性差，一般只适用于短期预测。时间序列预测一般反映三种实际变化规律：趋势变化、周期性变化、随机性变化。

在信息爆炸的时代，在分析相关问题时，有效信息的提取非常重要。主成分与因子分析一致，都是用来从大量信息中提取有效信息的方法。在数据信息提取模型上，EPS 知识服务平台提供了主成分分析、因子分析两种方法。由于这两种方法都是面对大量的数据信息，为此平台提供了一个预分析方法，以了解被分析数据的大致情况，如相关系数矩阵等。

EPS 知识服务平台考虑到很多用户需要对平台内变量进行多样化的操作，普通的变量转换无法完成这类需求，高级计算器就是为了满足变量的多元化计算需求而创造的。高级计算器以平台内的变量为计算单元，提供加减乘除、三角函数、对数转换、求余、求绝对值、开方等多样化的计算方法。

1.3 个人数据管理

EPS 知识服务平台为用户提供更多更加人性化的功能操作，如保存数据、上传本地数据、下载数据、分享数据等。拥有我的数据集与我的收藏夹。

2、智能报告生成平台

EPS 知识服务平台可以解决针对统计表格的数据分析需求，无论是平台自有的统计表格还是用户一键上传的统计表格，均能快速实现数据增删改、转置、排序选择等操作，提供折线、柱状、饼图等样式统计图表，针对不同结构的统计表单能够一键生成智能报告，图文并茂的对数据总体情况、分项情况进行全面剖析。

2.1 原始数据表格处理

用户可以直接在空白区域粘贴表格文件，也可以上传一个 excel 文件，系统后台会读取 excel 文件，将第一个表格数据显示在表格数据展示区域。并拥有修改单元格内容、合并单元格（向下合并，向右合并）、插入行、插入列、拆分行、拆分列、删除行、删除列、功能操作有：插入表格、删除表格等功能。

2.2 表格分析

基于处理过的原始数据表格，对原始表格进行指标、时间、单位、行列操作方面的处理。系统会自动的识别表格模式、统计地区、统计单位、统计年份、统计行业，将表格展示规范化。

2.3 智能报告生成

报告结构组要由三部分组成：封皮，目录，正文。报告正文内容分三部分：总体分析，分项分析，原始数据。总体分析是对表格的总体内容进行多角度描述。分项分析是对表格某个指标或者某个地区的数据进行多角度分析描述。原始数据是用户选中的被分析的数据范围。生产报告示例如下：

<p>各地区固定资产投资情况分析报告</p> <p>作者：张三</p> <p>时间：2016年6月12日</p>	<p>目录</p> <p>一、总体情况..... 3</p> <p>(一) 所有指标均呈增长态势..... 3</p> <p>(二) 施工项目持续快速增长..... 3</p> <p>(三) 新开工项目持续快速增长..... 4</p> <p>(四) 建成投产项目持续增长..... 4</p> <p>(五) 投资保持快速增长..... 5</p> <p>(六) 新增固定资产投资快速增长..... 5</p> <p>(七) 半数新开工项目集中在东部地区..... 6</p> <p>(八) 半数新开工项目集中在东部地区..... 9</p> <p>(九) 半数新建投产项目集中在东部地区..... 9</p> <p>(十) 近半数项目建成投产集中在东部地区..... 11</p> <p>(十一) 近半数的投资集中在东部地区..... 11</p> <p>(十二) 半数新增固定资产投资集中在东部地区..... 13</p> <p>(十三) 北京固定资产投资使用率最高..... 14</p> <p>二、分项情况..... 15</p> <p>(一) 东部地区..... 15</p> <p>1. 所有指标均呈增长态势..... 15</p> <p>2. 施工项目持续增长..... 16</p> <p>3. 新开工项目持续增长..... 17</p> <p>4. 建成投产项目持续增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅提高..... 17</p> <p>(二) 中部地区..... 17</p> <p>1. 所有指标均呈增长态势..... 17</p> <p>2. 施工项目持续快速增长..... 17</p> <p>3. 新开工项目持续快速增长..... 17</p> <p>4. 建成投产项目持续快速增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅提高..... 17</p> <p>(三) 西部地区..... 17</p> <p>1. 所有指标均呈增长态势..... 17</p> <p>2. 施工项目持续快速增长..... 17</p> <p>3. 新开工项目持续快速增长..... 17</p> <p>4. 建成投产项目持续快速增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅提高..... 17</p> <p>(四) 东北地区..... 17</p> <p>1. 所有指标均呈增长态势..... 17</p> <p>2. 施工项目持续快速增长..... 17</p> <p>3. 新开工项目持续快速增长..... 17</p> <p>4. 建成投产项目持续快速增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅提高..... 17</p> <p>(五) 北京..... 17</p> <p>1. 所有指标均呈增长态势..... 17</p> <p>2. 施工项目持续快速增长..... 17</p> <p>3. 新开工项目持续快速增长..... 17</p> <p>4. 建成投产项目持续快速增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅提高..... 17</p> <p>(六) 天津..... 17</p> <p>1. 所有指标均呈增长态势..... 17</p> <p>2. 施工项目持续快速增长..... 17</p> <p>3. 新开工项目持续快速增长..... 17</p> <p>4. 建成投产项目持续快速增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅提高..... 17</p>	<p>8. 固定资产投资使用率小幅下降..... 17</p> <p>(三) 西部地区..... 17</p> <p>1. 所有指标均呈增长态势..... 17</p> <p>2. 施工项目持续快速增长..... 17</p> <p>3. 新开工项目持续快速增长..... 17</p> <p>4. 建成投产项目持续快速增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅下降..... 17</p> <p>(四) 东北地区..... 17</p> <p>1. 所有指标均呈增长态势..... 17</p> <p>2. 施工项目持续快速增长..... 17</p> <p>3. 新开工项目持续快速增长..... 17</p> <p>4. 建成投产项目持续快速增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅下降..... 17</p> <p>(五) 北京..... 17</p> <p>1. 所有指标均呈增长态势..... 17</p> <p>2. 施工项目持续快速增长..... 17</p> <p>3. 新开工项目持续快速增长..... 17</p> <p>4. 建成投产项目持续快速增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅下降..... 17</p> <p>(六) 天津..... 17</p> <p>1. 所有指标均呈增长态势..... 17</p> <p>2. 施工项目持续快速增长..... 17</p> <p>3. 新开工项目持续快速增长..... 17</p> <p>4. 建成投产项目持续快速增长..... 17</p> <p>5. 项目建成投产率上半年基本持平..... 17</p> <p>6. 投资保持快速增长..... 17</p> <p>7. 新增固定资产投资快速增长..... 17</p> <p>8. 固定资产投资使用率小幅下降..... 17</p>
--	---	---



三、主题检索系统

EPS 知识服务平台，通过对文献的语义拆分与文献关键词的提取及挖掘分析，形成了庞大的主题词库，覆盖国内核心期刊实证类研究文献的全部研究主题。

1、知识发现平台

EPS 知识服务平台利用大数据技术，通过文本分析、机器学习、聚类关联、数据挖掘等技术手段，发现主题与统计指标间、不同主题间的关联关系，实现基于读者研究方向检索的指标推荐，发现海量实证研究文献中所凝结的作者的知识与智慧。

2、文献计量分析

EPS 知识服务平台基于共现关联统计理论研发了文件计量分析平台，通过对海量文献元数据的分类汇总、数据清洗、共现计算、关联分析，可以提供基于元数据的研究趋势、热点分布、知识图谱等功能的文献计量分析服务，便于用户发现并理解大量文献背后隐藏的信息，解决文献超载问题，为知识发现提供更多可能。

2.1 研究趋势

为了判断某研究方向/研究主题是否还具有研究价值，用户往往需要阅读大量的原始论文，很容易迷失在浩如烟海的论文之中，而经历海量阅读之后，用户依旧很难断定，自己选择的研究方向是否还具备研究价值。为了解决这个问题，文献计量分析平台研发了研究趋势功能。

研究趋势通过对用户检索得出的批量最相关文献进行多角度挖掘抽取，告诉用户基于某关键词的研究趋势。主要的趋势指标有 6 个，分别是论文数量、收录期刊数量、作者数量、学科数量、机构数量、基金数量，对于每个指标又有两个细项：每年数量和累积数量。

2.2 研究热点

研究热点通过对用户检索得出的批量最相关文献进行多角度挖掘抽取，告诉用户当前被检索的研究主题的热点分布。主要的热点指标有 5 个：热点期刊、热点作者、热点学科、热点机构、热点基金，与研究趋势一样，对于每个热点同样从两个角度统计：单年和累积。

2.3 热点期刊

主要是登载与研究主题相关论文的前 20 的期刊列表以及每个期刊登载的相关论文的数量；累年登载与研究主题相关论文数量前 20 的期刊及每个期刊登载的论文数量。

2.4 热点学者

主要是当年发表与研究主题相关论文数量前 20 的作者列表及每个作者发表论文的数量；累年发表与研究主题相关论文数量前 20 的作者列表以及每个作者发表的论文数量。

2.5 热点学科

主要是当年与研究主题相关论文数量前 20 的学科列表及每个学科发表论文的数量；累年发表与研究主题相关论文数量前 20 的学科列表以及每个学科发表的论文数量。

2.6 热点机构

主要是当年发表与研究主题相关论文数量前 20 的机构列表及每个机构发表论文的数量；累年发表与研究主题相关论文数量前 20 的机构列表以及每个机构发表的论文数量。

2.7 热点基金

主要是当年支持与研究主题相关论文数量前 20 的基金列表及每个基金支持发表论文的数量；累年支持与研究主题相关论文数量前 20 的基金列表以及每个基金支持发表的论文数量。

3、知识图谱

EPS 知识服务平台通过信息可视化技术描述知识资源的相关关系，显示知识发展进程与核心结构，通过知识图谱帮助用户获取最需要的信息，使用户的搜索具有深度与广度，为用户构建与搜索结果相关的完整知识体系，并帮助用户发现某个新的事实或新的联系。

3.1 组合衍生词图谱

组合衍生词是指基于某个研究主题演进而来的新词，很有可能代表研究主题的细化，以“创新”为例，“技术创新”就是创新的组合衍生词。组合衍生词的数量变化是衡量研究演进和扩散速度的一个很好指标。

3.2 机构合作图谱

大部分研究人员认为数据记录中作者合作就说明机构存在合作，只要论文的作者属于不同的研究机构就认为该论文是这些不同机构合作的结果。在社会网络分析方法中，如果一个节点处于许多其他点对的最短途径上，就说该点具有较高的中介中心度，它可能起到很重要的“中介”作用，处于网络的中心。

学术机构知识图谱是以促进学术机构学科发展和加强科研合作目的，利用科学计量学、计算机科学、社会网络分析和复杂网络等理论和方法，对学术机构知识载体信息进行绘制、挖掘、分析，显示其所创造的知识 and 知识之间的关系。

3.3 机构+关键词图谱

机构+关键词知识图谱是以分析学术机构学科发展现状和评估未来发展方向为目的，利用科学计量学、计算机科学、社会网络分析和复杂网络等理论和方法，对学术机构的研究方向、研究主题信息进行绘制、挖掘、分析，显示其所创造的知识 and 知识之间的关系。

3.4 共现关键词图谱

如果说关键词代表的是一个科学概念或科研领域，共现关系则是进一步揭示了其科学命题或科研方向。

3.5 作者合作图谱（显性）

合作研究是实现资源优势互补，促进知识交流和共享的重要方式，在大科学时代，面对一些跨学科、跨地域、高难度、大规模的复杂问题，科研合作尤为重要。科学计量学家在研究中甚至还发现：科学合作不但能够促进科学家的成果产出能力，而且能提高研究成果的质量和影响力；合著论文的著者人数与论文产生的影响因子呈正相关；合著论文比独著论文更容易被期刊接受并能产生较大影响；合作群体越大，研究成果越具有影响力。

所谓作者显性合著，是基于单篇论文提出的，如果两个作者同时出现在一篇论文中，称之为显性合著。在作者显性合著图谱中，DFINDER 基于批量检索结果抽取显性合作网路，用户可以通过操作年份实现单年合著学者和累积合著学者网络的可视化。通过点击作者，可以查看该作者发表的所有论文，通过点击合著连线，可以查看合著相关论文列表。

3.6 作者合作图谱（隐性）

隐性的作者合作是指那些具有相同研究方向（研究领域、研究方法），使用相同关键词而形成的一种新的隐藏的关系网络，在这个网络之中，相关研究者之间可能是完全不认识的。这种隐性的作者合作网络分析也具有他的研究意义，比如对研究者如何寻找合作伙伴可以提供帮助。

EPS 知识服务平台首先基于检索得出的批量结果提取出排名靠前的高频关键词，接着围绕某个关键词领域的提取本研究领域的作者，绘制作者隐性合作图谱。