

数字资源知识共享与知识再利用 模式与方法研究

中国科学技术信息研究所

刘耀

2022年1月

内容提要



一、研究介绍



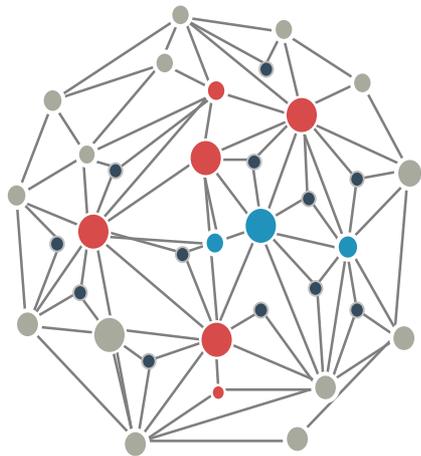
二、关键技术



三、智能服务平台及应用研究



四、总结与展望



/01

一、研究介绍

研究动态及问题

解决方案

理论方法

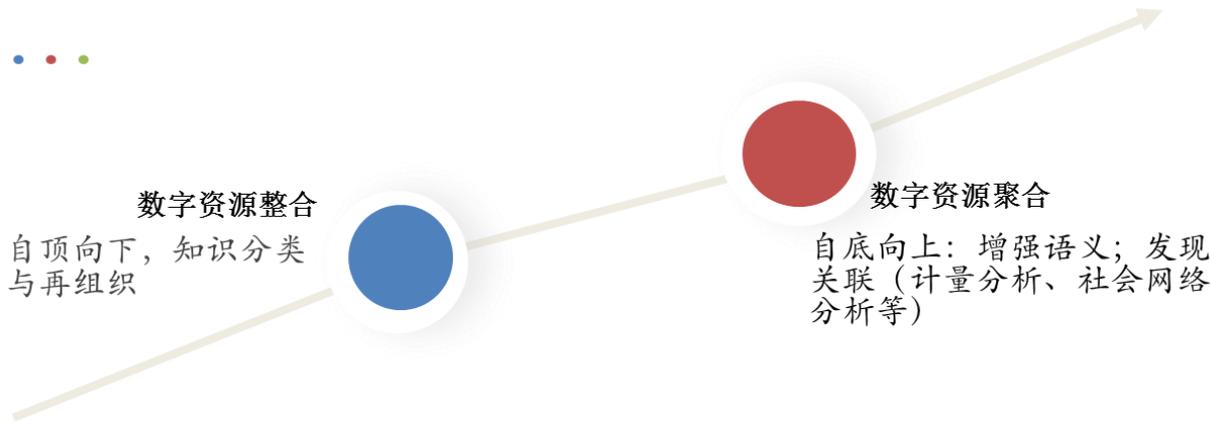
研究重点

研究动态

(1) 知识组织模式研究

知识组织模式经历了从数字资源整合向数字资源聚合的转变。

- **问题1**：现存知识组织模式研究项目主要研究领域是语言学、图书馆和情报学，存在领域限定性强、工程量大、周期长等等问题，在工程中应用灵活性较差。



研究动态

(2) 知识利用模式研究

当前主流知识利用模式包括学科信息门户服务模式、专业化知识服务模式、个性化定制知识服务模式、数字化参考咨询服务模式和自助式知识服务模式。

问题2：针对知识服务的研究主要是对服务形式的分析，未能很好解决业务驱动下的知识组织与利用的自动化问题。



解决方案

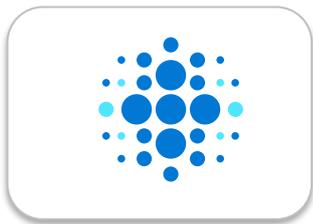
现存知识服务两大问题的解决需要从业务需求入手，以**业务组织模型理论与方法**为指导，通过**业务模型自动构建研究**，形成个性化业务模型；针对当前高昂成本构建的知识库中数据利用程度过低的问题，利用**单篇全解析及多篇路径自组织**等技术手段，为真实项目的应用服务提供技术支持；以**新型服务平台及具体服务**为示范，解决知识体系构建与重用的现实问题。



理论研究

业务组织模型的理论与方法研究：

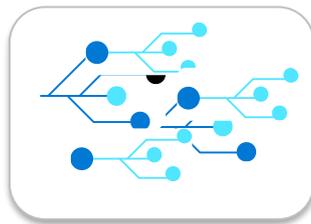
通过对业务问题的分解、业务链条单元的描述、业务流程的结构化能力，建立业务问题-知识事件的关联模型，将业务单元与知识载体相匹配，构建业务组织模型。



业务问题分解



业务链条单元描述



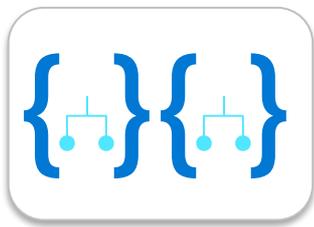
业务流程的结构化

能力建设

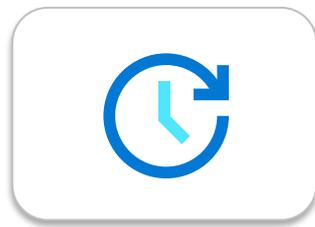
本研究包含以下两个重点能力建设：

(1) 业务模型的快速构建：解决业务知识组织与知识再利用之间存在断层，当前知识服务缺少业务驱动能力的问题。

(2) 快速解析资源的技术能力：解决目前领域知识体系构建成本过高的问题。



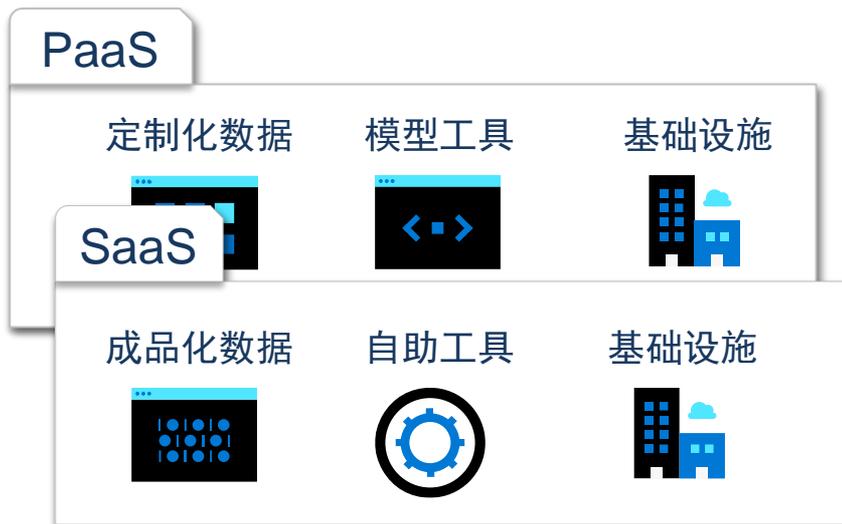
业务模型的快速构建

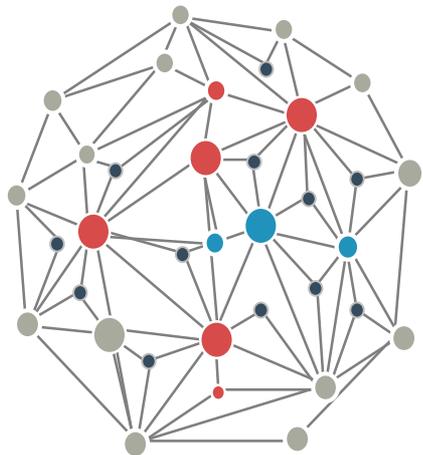


快速解析资源

新型服务机制

为了解决知识组织完成后的知识再利用问题，基于以上理论研究和能力建设，本研究提出一个新型的服务模式：**基于PaaS以及SaaS的新型服务模式**





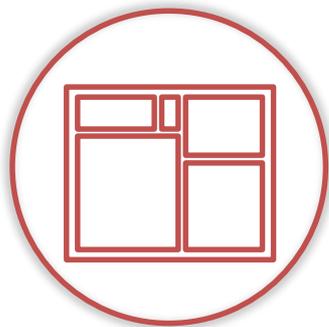
/02

二、关键技术

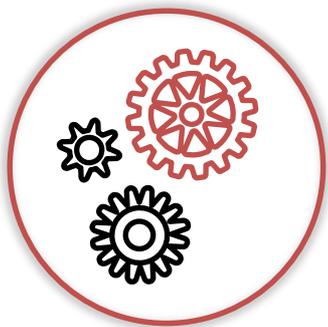
1. 业务自动成与业务复用研究
 2. 敏捷的专业知识组织方法研究
-

1. 业务自动生成与业务复用研究

业务自动生成与业务复用研究包括以下几个研究内容：



业务模型
自动构建



单业务资源
解析



多系统知
识关联



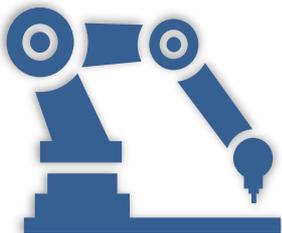
业务资源再利
用路径自组织

1. 业务自动生成与业务复用研究

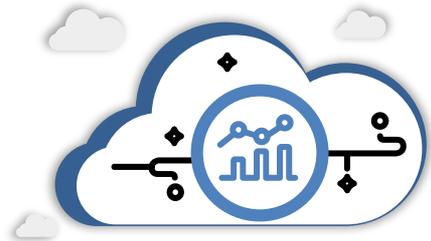
➤ 业务模型自动构建



业务问题自动化分析，构建面向需求的个性化业务资源库



围绕业务资源研究知识的定义及抽取



广泛关联资源，对业务知识进行再利用

业务问题驱动下的业务模型自动生成

从业务资源抽取知识的关键技术

认知指导下的业务流程自动生成

1. 业务自动生成与业务复用研究

➤ 单业务知识资源库

1) 知

2) 知

最后，

构形

果构建单业务知识资源库。

Request-Handler (qt)
/select
resourceType: event.xsd
start, rows: 0, 10
wt: json
indent

单业务知识资源库：
以新闻解析业务为例

结构化处理

结果网页列表

```

request.getParameter("${label}");

class="bads_global-header-text"</
header__authors f-ui"</header>
header"</header>

-nav</navbar>
-simple-tabs"</navbar>
s="m-quick-links"</navbar>

-body" -attr- data-behavior="Handle
ndemic]</articleTitle>

v/article/header/n1 -attr- class="m-article-header__title f-h1f

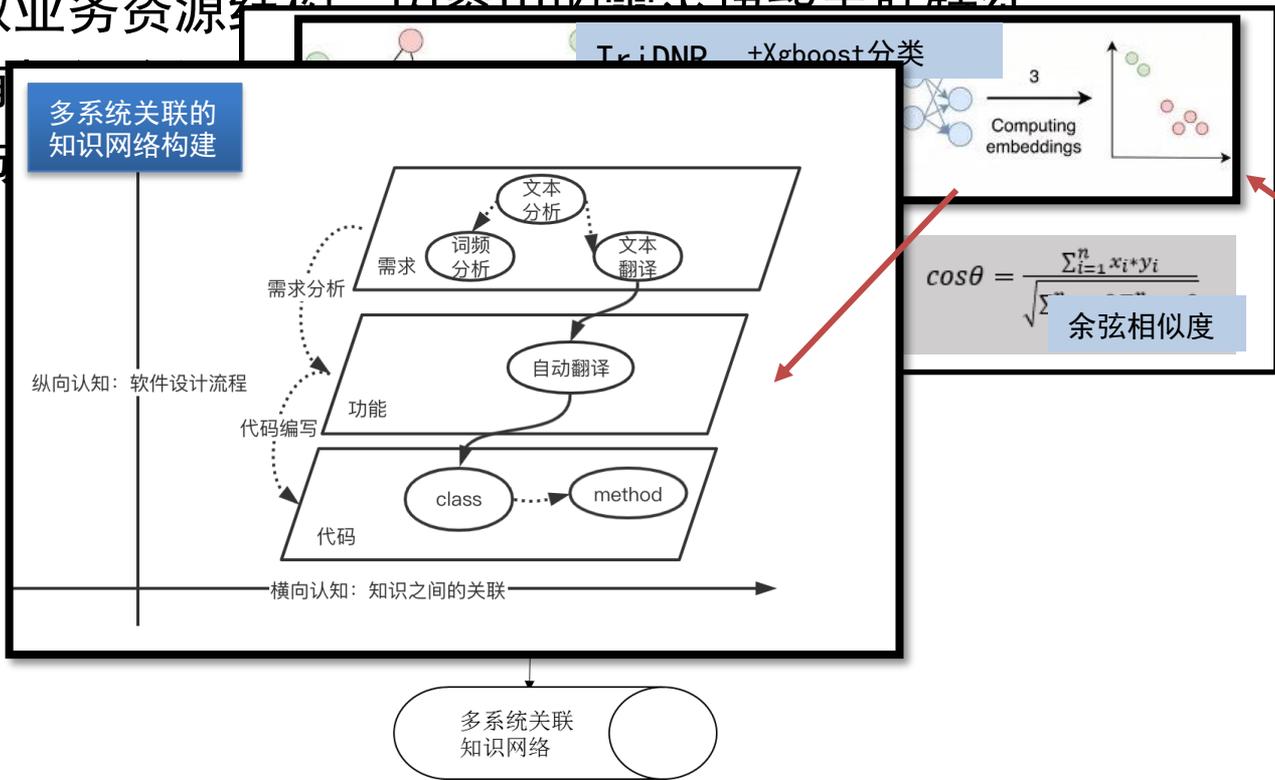
<articleDate>Dec 3, 2020</articleDate>
<articleDateTag>html/body/div/div/main/div/div/div/div/div/article/div/div/div/aside/ul/li/a/span/span -attr- class="m-s
                    
```

1. 业务自动生成与业务复用研究

➤ 多系统知识关联

1) 分析提取业务资源结构 内容中的需求功能关联特征

2) 关联具有
接关系，形成
供基础。



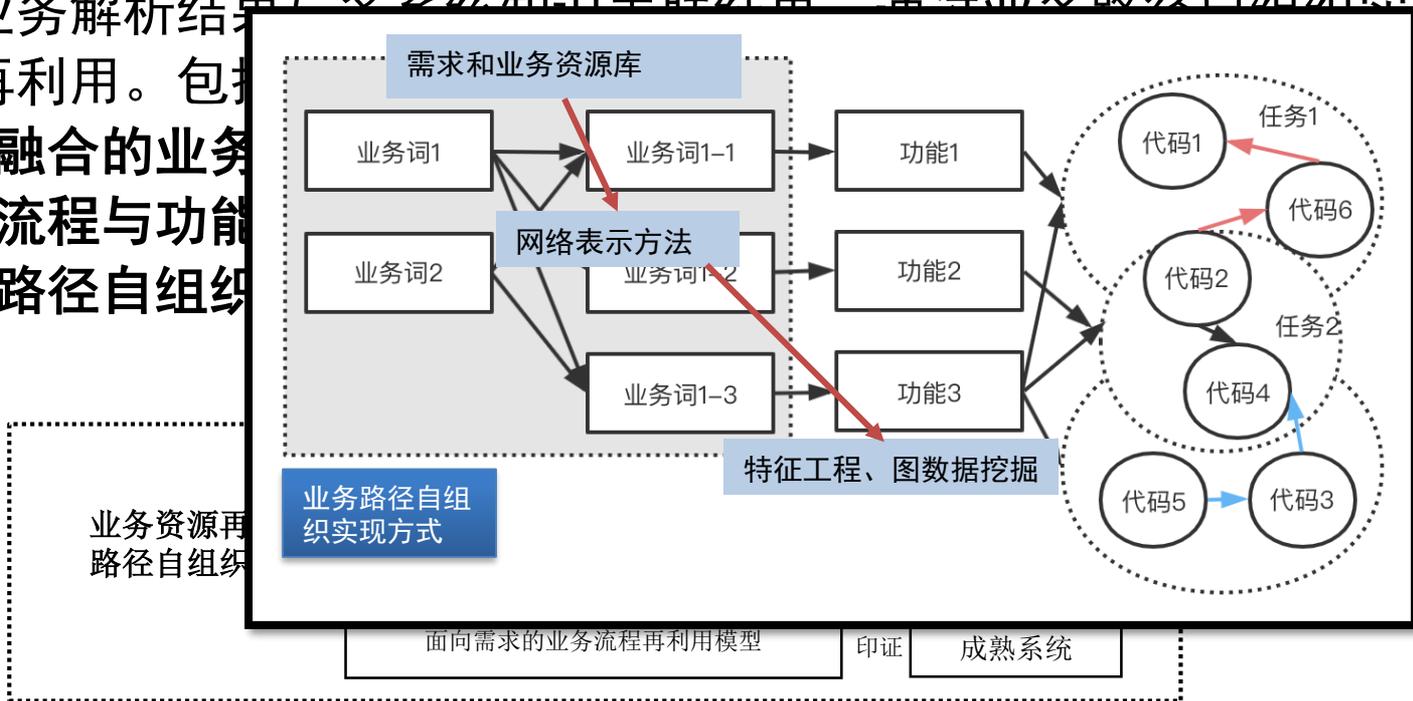
点间的链
关系提

1. 业务自动生成与业务复用研究

➤ 业务资源再利用路径自组织

基于单业务解析结果与多系统知识关联应用，通过业务路径自组织实现业务资源再利用。包括

- 1) 多源融合的业务
- 2) 业务流程与功能
- 3) 业务路径自组织



2 敏捷的专业知识组织方法研究

敏捷专业知识组织方法研究通过研究复杂语境下的知识发现、开放关系中的知识发现和知识映射的路径发现实现单篇文本知识结构全解析。在此基础上，数字资源知识自组织模型及关键技术主要解决多篇文本中知识自组织模型的构建与路径识别问题。

2 敏捷的专业知识组织方法研究

➤ 单篇文本知识结

(1) 复杂语境下的

不进行显性标注的情
语义单元；结合形式
容语义进行先验知识
及实体到实体的解码

(2) 开放关系中的

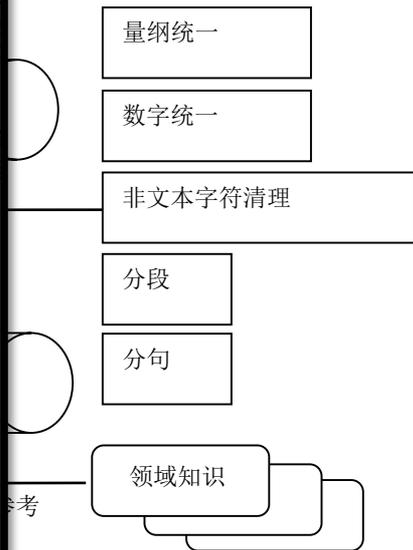
通过先验知识与文本
映射进行知识发现、
义的融合，实现文本
与深度结构化。

```

<allStructure>
  <source>两都医案</source>
  <content>姚文学康伯，何相国客也。长夏时，自云虚病，弱不能支，饮食强不能进，
  薏为君，厚朴、白藕豆、广陈皮、白茯苓、炙甘草、石斛为臣，山楂、麦芽为佐，枳壳
  脉不显，而沉滑尤甚。若不急治标、缓治本，标病方炽，岂能以静养治本乎？气血冲和
  痰滞而何哉？因决意用琥珀丹一粒，痰滞尽下，遂觉霍然无恙。后用六君、四物剂，数
  <splitwords>姚\文学\康伯\，\何\相国\客\也\。\\长夏\时\，\自\云\虚病\，\弱\
  \则\脉沉滑\，\是\不可\作\虚损\治\，\须\先\以\清暑\舒郁\为主\，\脾胃\冲和\，
  分\，\又\煎\加\姜汁\一\茶匙\为\引\。\\二\服\，\顿觉\神清气爽\，\胸中\灵快\。
  \，\标\病\方\炽\，\岂能\以\静养\治本\乎\？\气\血\冲和\，\万病\不\生\，\一\有
  \右\厢\有\块\觉\痛\。\\余\曰\：\经\云\无\滞\不\作\痛\，\非\痰滞\而\何\哉\？\因
  决\作\痛\。\\一\丸\不\利\，\再\一\丸\，\以\利\为\效\，\治\痢疾\亦\妙\。\\凡\一切\
  黄芩\五\钱\积实\五\钱\山查\肉\五\钱\九\蒸\九\晒\大\黄\一\两\五\钱\共\为\细末\
  <concepts>
    <conceptName>人名</conceptName>
    <conceptSent>姚文学康伯</conceptSent>
  </concepts>
  <concepts>
    <conceptName>无关</conceptName>
    <conceptSent>何相国客也</conceptSent>
  </concepts>
  <concepts>
    <conceptName>时间</conceptName>
    <conceptSent>长夏时</conceptSent>
  </concepts>
  <concepts>
    <conceptName>病症</conceptName>
    <conceptSent>自云虚病</conceptSent>
  </concepts>
    
```

单篇文本全解
析结果示例

流程图

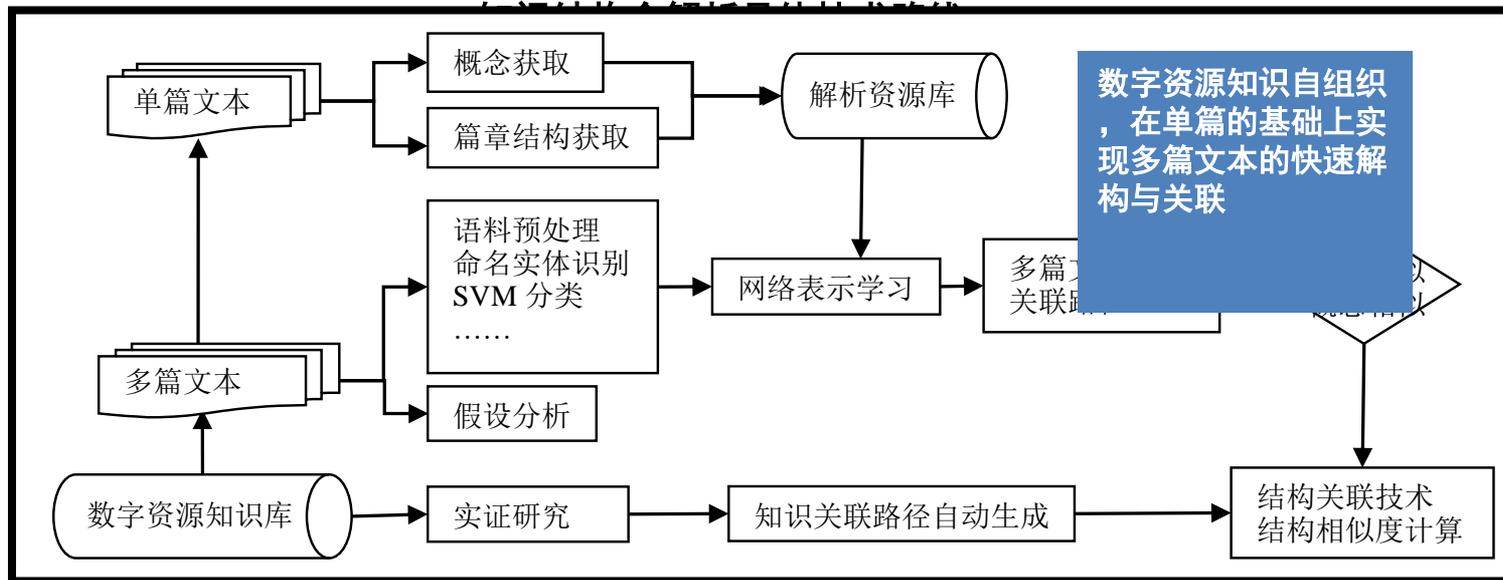


2 敏捷的专业知识组织方法研究

➤ 单篇文本知识结构全解析关键技术

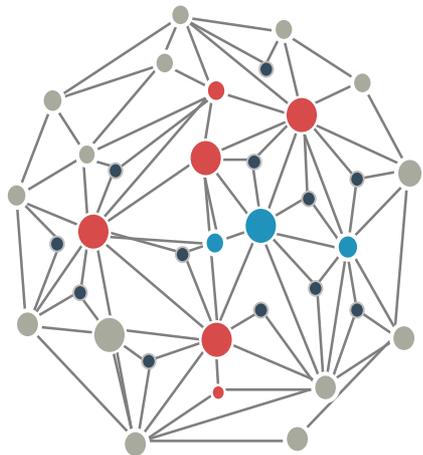
(3) 知识映射路径发现：

用大量语料进行任意领域知识体系的发现，将所发现的知识体系进行精炼，指导单篇文本知识结构生成，从而实现全解析的目标。



小结

针对业务知识组织与知识再利用之间存在的断层问题，本节展示了**业务自动生成与业务复用技术**，针对当前知识组织模式中缺少灵活构筑资源的个性化解决方案、且高昂成本构建的知识库中数据利用程度过低的问题，本节展示了**高效敏捷的知识组织机制**，为真实项目的应用服务提供技术支持。



/03

三、智能服务平台及应用示范研究

3.1 SaaS服务模式

3.2 PaaS服务模式

3. 智能服务平台及应用示范研究

为了解决知识组织完成后的知识再利用问题，需要实现多资源聚合与多场景融合，建成标准化、组件化、可扩展、可灵活动态构建的服务平台。包括：

- (1) 成品化数据+自助工具的SaaS服务模式。
- (2) 定制化数据+模型工具的PaaS服务模式。

3.1 服务模式探索：SaaS服务模式

SaaS：软件即服务。

作用：将应用作为服务提供给客户。

优势：通过SaaS这种模式，用户只要接上网络并通过浏览器，就能直接使用在云端上运行的应用。消除了企业购买、构建和维护基础设施和应用程序的需要。

人群：SaaS主要面对的是**普通用户**。



泛专题自动生成

泛专题作为SaaS服务模式服务示范，本质上是一定数量级的个体专题集合，强调对单个数字资源挖掘的全面与深入、多个数字资源之间的广泛知识关联。包括：

- (1) 知识网络构建与网络表示。
- (2) 多样化排序与展示。

泛专题自动生成——SaaS服务模式应用实例

首先用户联网登录现有平台，进入主页，选择泛专题功能：



泛专题自动生成——SaaS服务模式应用实例

中医古籍文本知识图谱构建与应用系统

首页 古籍全览 知识检索 科研服务 泛专题 行业动态 关于我们

当前位置: 首页 > 科研服务 > 泛专题

+ 添加

名称	简介	操作
测试	测试	使用模板 删除模板
ces2		使用模板 删除模板
测试2		
药食同源	以所有可作为食物的药物组成为方剂	
四君子汤相关药食专题	四君子	
十枣汤		

用户根据业务需要，自行添加模板或使用现有模板。用户一切操作都在现有网站上执行，无需管理底层设施。使用的数据是成品数据。

泛专题自动生成——SaaS服务模式应用实例

中医古籍文本知识图谱构建与应用系统 首页 古籍全览 知识检索 科研服务 泛专题 行业动态 关于我们

当前位置：首页> 科研服务> 泛专题

+ 添加

名称	简介	操作
测试	测试	使用模板 删除模板
ces2		使用模板 删除模板
测试2		使用模板 删除模板
药食同源	以所有可作为食物的药物组成为方剂	
四君子汤相关药食专题	四君子	
十枣汤		

选择模板，查看平台生成的药食同源泛专题集合。同时也可查看单个知识的详细介绍。

< 1 >

3.2 服务模式探索：PaaS服务模式

PaaS 平台即服务。

作用：将开发平台作为服务提供给用户。

人群：PaaS主要的用户是**开发人员**。

优势：通过PaaS这种模式，用户可以在一个包括SDK，文档和测试环境等在内的开发平台上非常方便地编写应用，而且不论是在部署，或者在运行的时候，用户都无需为服务器，操作系统，网络和存储等资源的管理操心，这些繁琐的工作都由PaaS供应商负责处理。

PaaS服务模式:

PaaS的服务模式允许用户根据现有平台的能力完成业务重组，实现从数据到服务的全新业务。

使用统一认证登录进入开发平台

使用爬虫平台获取资源

使用或上传数据模型至数据交换池

使用或上传工具至数据处理工具集

调用或上传算法至算法平台

使用或上传前台组件

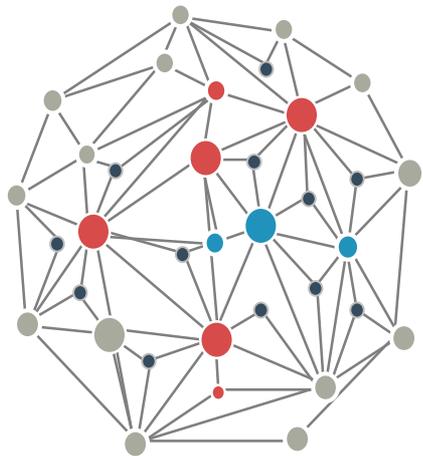
通过算法注册)或引用查询)。



The screenshot displays a development environment with two windows. The left window shows a component library with categories like 'carousel', 'chart', 'force-hotword', 'layout', and 'types-resource-list'. The right window shows a code editor with a JavaScript component definition for a '通用柱状图' (General Bar Chart). A red arrow points from the 'chart' component in the library to the corresponding code in the editor.

```
1 <script id="通用柱状图-6d4ee5f9-6b01-42e0-bbe6-c6045407e1c">
2 {
3   let $component = {name: null,init: null,option: null,ajax: null}
4   $component.option.resourceTitle = '通用柱状图'
5   $component.option.statisticsField = 'regenerateReport_info_classification'
6   let base = '/informationAnalysis'
7   $component.ajax = () => {
8     return $http.post(base + '/policy_visualizationByOption.action',{
9       filter:'type:' + $component.option.resourceTitle,
10      option: $component.option.statisticsField,
11      graphType: 'bar'
12    })
13   }
14   $component.initialize = () => {
15     $component.$chart = echarts.init(document.getElementById('chartbar-6d4ee5f9-6b01-42e0-bbe6-c6045407e1c'))
16   }
17   $component.name = '通用柱状图'
18   $component.render = result => {
19     if(result.result) result = JSON.parse(result.data)
20     let option = {
21       xAxis: {
22         type: 'category',
23         data: []
24       },
25       yAxis: {
26         type: 'value'
27       },
28       series: [{
29         data: [],
30         type: 'bar',
31       }
32     ]
33     option.xAxis.data = result.xAxis
34     option.series[0].data=result.series[0].data
35     $component.$chart.setOption(option)
36   }
37   if($component.initialize) $component.initialize()
38   $component.ajax().then(res=>
```

使用或上传前台组件，完成数据的可视化展示



/04

总结与展望

总结

基于技术的新型知识共享模式：基于扎实的技术与工程能力，形成新型知识共享与服务模式，避免纸上谈兵的知识体系构建研究。

快速灵活的业务模型构建技术：通过单篇文本全解析与路径自组织技术研究，实现高精度的资源解析及知识体系构建。避免了大型领域知识库构建的高昂成本，为资源共享与再利用赋能。

谢谢！
