

Platform for Applying, Researching and Developing Intelligent Search Engine (PARADISE)

闫宏飞

<http://sewm.pku.edu.cn>

September 16, 2007

导航

1 项目由来

2 PARADISE

项目任务

课题是由一个平台，四个任务组成的，达到平台能够展示四个任务

平台是：

- ① 一个开放式智能化中文搜索引擎平台

任务：

- ① 构建适用于中文互联网的文本内容抽取系统及网页文本内容结构化分析
- ② 基于内容的多媒体信息检索
- ③ 基于自然语言理解的查询分析
- ④ 基于用户属性挖掘的个性化检索

Projects and Staff of SEWM

	YHF	PB	XZM	YCL	HLE	CC	DN	HJ	LY	TQC	GSL	YZF	DMC	SDD	YR	ZDS
Q&A	√			.4		.2		.2								
.RAS\$	√					.1								√		√
^EAER				.6			1.0						1.0			
★PARA	√	√						.6	√	√		√				
.SCUT\$			√													
^DASI		√														
DIMCID						.7										
^ODIR			√		√											
BIG	√													√		√
.WESSI			√								√					
RGC			√								√					
.IRICH		√						.2	√							

★ denotes the most important project, ^ denotes beginning, \$ denotes ending, . denotes first applier.

<http://sewm.pku.edu.cn/about.html#projects>

EAER=李Nsf2,PARADISE=863₁,DASI=民航,WESSI=李Nsf1,RAS=闫Nsf,ODIR=863₂,BIG=闫863,DIMCID=支撑

后备队

- 李进, 毛先领(35hrs); 朱磊(10-15hrs), 樊锴(16-20hrs), 寿思聪(20hrs), 陈日閃; 李庚, 高泓YU, 于勇建

指导教师与学生

- LXM: HLE, CC, YCL, DN, HJ, LY, GSL, LG, DMC
- YHF: SDD, ZDS
- PB: TQC, YZF

Picasa Web Albums

[My Photos](#) | [My Favorites](#) | [My Public Gallery](#)

Community Photos | Search Photos



paradise20070727brainstorm
Photos 10 - 2 MB
Aug 14, 2007
Public
[Edit Album Properties](#)
[Change Album Cover](#)

[Upload Photos](#)
[Embed Slideshow](#)
[Link to this Album](#) ▶
[Delete Album](#)



Yan

SlideShow | View Map | Organize | Edit Captions | Order Prints | Share Album | Size: [] [] []



http://picasaweb.google.com/yhf102R/Paradise20070814

Gmail - Inbox | Index of /cs101 | Google Docs | preseat的文档... | index的文档... | PARADISEFrame... | Paradise | TianwangGForge... | Picasa Web ...

paradise20070814
 Photos: 16 - 4 MB
 Aug 14, 2007
 Public
[Edit Album Properties](#)
[Change Album Cover](#)

[Upload Photos](#)
[Embed Slideshow](#)
[Link to this Album](#) ▶
[Delete Album](#)

Yan

查找: dyf

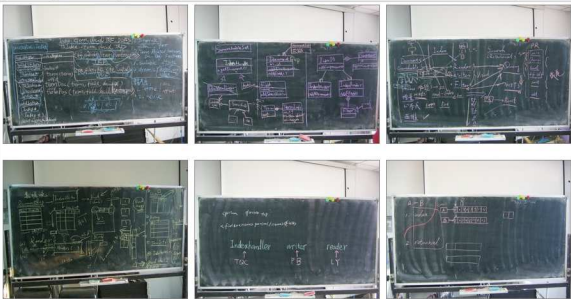


paradiseIndexDesign
Photos 6 - 1 MB
Sep 1, 2007
Public
[Edit Album Properties](#)
[Change Album Cover](#)

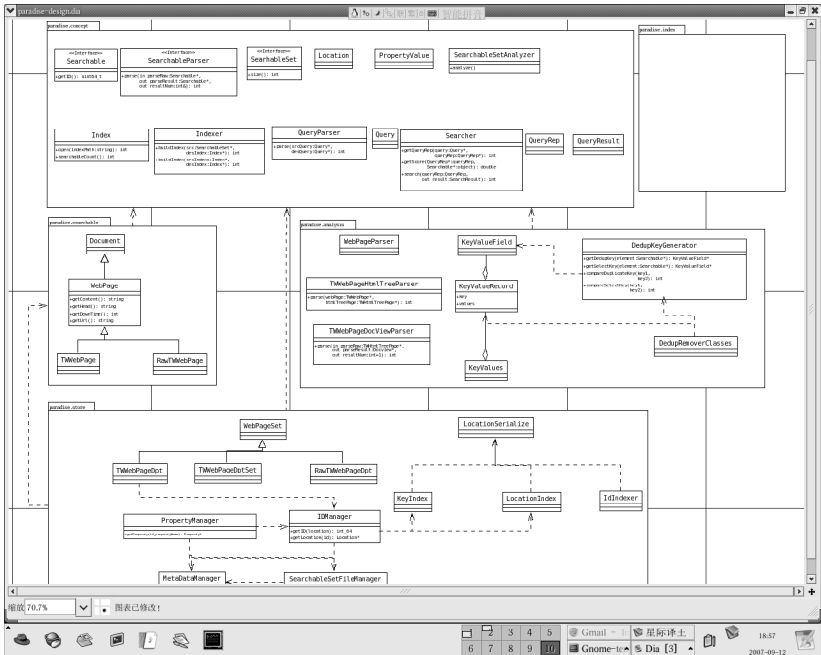
[Upload Photos](#)
[Embed Slideshow](#)
[Link to this Album](#)
[Delete Album](#)



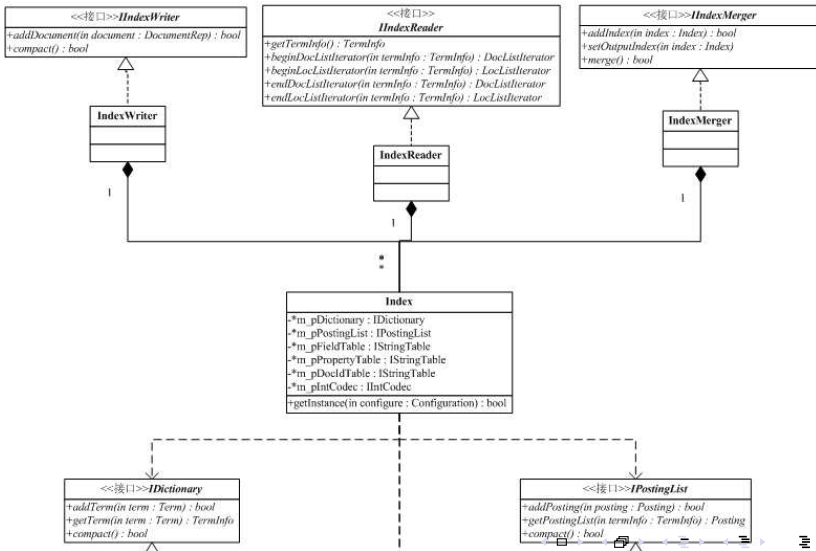
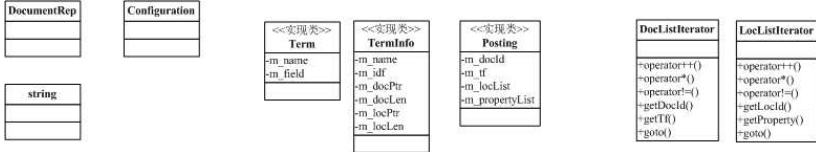
Yan



RSS



缩放 70.7% 图表已修改!



Paradise 文件列表

组合类型列表
 类继承关系
 组合类型成员
 类继承关系图
 名字空间列表
 名字空间成员
 文件列表

- analysis/include/DOCParser.hpp
- analysis/include/PDFParser.hpp
- analysis/include/ReduplicateRemover.hpp
- analysis/include/SearchableParser.hpp
- analysis/include/SearchableSetAnalyzer.hpp
- analysis/include/Token.hpp
- analysis/include/TokenStream.hpp
- analysis/include/TWWebPageDocViewParser.hpp
- analysis/include/TWWebPageHtmlTreeParser.hpp
- analysis/include/WebPageParser.hpp
- analysis/src/DOCParser.cpp
- analysis/src/PDFParser.cpp
- analysis/src/ReduplicateRemover.cpp
- analysis/src/TWWebPageDocViewParser.cpp
- analysis/src/TWWebPageHtmlTreeParser.cpp
- analysis/test/pseudoPageRankGen.cpp
- analysis/test/DOCParser.cpp
- analysis/test/TWParse.cpp
- analysis/test/TPDFParser.cpp
- analysis/test/TWParse.cpp
- analysis/test/TWParser.cpp
- analysis/twparser/include/ChineseTokenizer.hpp
- analysis/twparser/include/HtmlTree.hpp
- analysis/twparser/include/HtmlTreeParser.hpp
- analysis/twparser/include/HtmlTreePlus.hpp
- analysis/twparser/include/HtmlTreeParserPlus.hpp
- analysis/twparser/include/JudgeType.hpp
- analysis/twparser/include/String.hpp
- analysis/twparser/include/Tokenizer.hpp
- analysis/twparser/include/TokenTag.hpp
- analysis/twparser/include/TWDocView.hpp
- analysis/twparser/include/TWMD5.hpp
- analysis/twparser/include/TWPageCleaner.hpp
- analysis/twparser/include/TWParser.hpp
- analysis/twparser/include/TWParserParser.hpp

Paradise 文件列表

这里列出所有文件，附带简要说明：

analysis/include/DOCParser.hpp [代码]	Definition of class <i>CDOCParser</i>
analysis/include/PDFParser.hpp [代码]	Definition of class <i>CPDFParser</i>
analysis/include/ReduplicateRemover.hpp [代码]	
analysis/include/SearchableParser.hpp [代码]	
analysis/include/SearchableSetAnalyzer.hpp [代码]	
analysis/include/Token.hpp [代码]	
analysis/include/TokenStream.hpp [代码]	
analysis/include/TWWebPageDocViewParser.hpp [代码]	
analysis/include/TWWebPageHtmlTreeParser.hpp [代码]	
analysis/include/WebPageParser.hpp [代码]	
analysis/src/DOCParser.cpp [代码]	Implementation of <i>CDOCParser</i>
analysis/src/PDFParser.cpp [代码]	Implementation of class <i>CPDFParser</i>
analysis/src/ReduplicateRemover.cpp [代码]	
analysis/src/TWWebPageDocViewParser.cpp [代码]	
analysis/src/TWWebPageHtmlTreeParser.cpp [代码]	
analysis/test/pseudoPageRankGen.cpp [代码]	
analysis/test/DOCParser.cpp [代码]	Test program of class <i>CDOCParser</i>
analysis/test/TWParse.cpp [代码]	
analysis/test/TPDFParser.cpp [代码]	Test program of class <i>CPDFParser</i>
analysis/test/TWParse.cpp [代码]	
analysis/twparser/include/ChineseTokenizer.hpp [代码]	
analysis/twparser/include/HtmlTree.hpp [代码]	A <i>Html Tree Parser</i>
analysis/twparser/include/HtmlTreeParser.hpp [代码]	
analysis/twparser/include/HtmlTreePlus.hpp [代码]	
analysis/twparser/include/HtmlTreeParserPlus.hpp [代码]	
analysis/twparser/include/JudgeType.hpp [代码]	Get the type of a page
analysis/twparser/include/String.hpp [代码]	
analysis/twparser/include/Tokenizer.hpp [代码]	
analysis/twparser/include/TokenTag.hpp [代码]	
analysis/twparser/include/TWDocView.hpp [代码]	Redefine of <i>Tianwang Docview</i> , which raise by <i>Zhangzhigang</i>
analysis/twparser/include/TWMD5.hpp [代码]	
analysis/twparser/include/TWPageCleaner.hpp [代码]	For clean the page and return a clean content different clean method to different Type of page.
analysis/twparser/include/TWParser.hpp [代码]	Facade of <i>Tianwang Docview</i> algorithm, note: <i>htmltree</i> use <i>Tidy</i>
analysis/twparser/include/TWParserParser.hpp [代码]	Main algorithm of <i>Docview</i> , which is invent by <i>Zhang Zhigang</i>

查找: dyl

http://tianwang.gforge.cn/

Gmail - Inbox Index of /cs101 Google Docs preatext的文档... index的文档... PARADISEFrame... Paradise TianwangGForge...

FORGE

软件/群组 [] 搜索 登录 新帐号

首页 我的首页 软件地图 分享代码 目录结构

欢迎来到天网GForge平台。

天网GForge是一个软件开发管理平台。它包含论坛，邮件列表用于开发者之间的协作。它提供方便的源代码管理（CVS和Subversion）。当你新建一个项目时，GForge根据你的设置自动生成一个源代码库。

其他工具：

- 文件发布管理
- 文档管理
- 新闻发布
- 用户和管理员状态
- 任务管理

本系统的数据库每天进行备份，以保证您的数据不会丢失。

最新消息

天网seplatform启动了
Zhiheng YANG - 2007-07-18 09:37 - seplatform
2007年7月2日，天网seplatform正式启动了。
(0 评论) [阅读 更多/评论]

项目文档在线浏览
彭波 - 2007-07-18 09:37 - TPlatform
现在用google docs来协同编辑文档了。
(0 评论) [阅读 更多/评论]

最新进展
李合 - 2006-12-19 20:19 - HEINIU Code Digger
PiDan完成了源代码聚类分析Package，使用了CCFinder代码卷检测工具。
(0 评论) [阅读 更多/评论]

加快速度！
wei cao - 2006-12-19 20:19 - HEINIU Code Digger
code search project计划
milestone 1.0目标
(1 评论) [阅读 更多/评论]

强烈庆祝code search的milestone 1.0顺利完成！！
wei cao - 2006-12-19 20:18 - HEINIU Code Digger
后台和前台已经联通
当前测试数据量包括一百个java文件
用tomcat了web页面
(0 评论) [阅读 更多/评论]

- 本项目活力充足！ 2006-12-11 21:53
- 项目开始啦！ 2006-12-05 15:08
- 终于确定做什么了 2006-11-29 20:17
- GForge平台复活了！ 2006-11-28 02:12
- snke 0.6.2发布 2006-10-13 09:47

[新闻存档]

欢迎使用

本服务目前主要针对校园网内用户（虽然外网用户也可以访问，不过我们不保证外网用户的可用性）提供服务。出于安全的考虑，CVS代码库通过peerover访问的方式被关闭；项目成员仍可通过ssh方式访问CVS。我们强烈推荐您使用svn进行源代码的版本控制，它更安全而且用户界面更友好。

要使用本系统的服务，您必须先注册。通过右上角的链接您可以进行注册和登录。

新手帮助

请先查看这个手册。关于gforge使用的讨论请到这个公共论坛，各种问题可以在这里讨论，你的问题或许马上在这里可以找到答案。如果你确信是系统发生什么故障了，可以给我发信。

TianwangGForge 统计资料

登录项目: 48
注册用户: 81

项目下载排行榜

(169) 基于软件的元搜索引擎
(5) liuyuan个人SVN实验场
[更多]

用户评价排行榜

没有统计信息

本周活力最强排行榜

(100.0%) TPlatform
(75.0%) cs101 fall 2007
(50.0%) mini search engine
(25.0%) 基于软件的元搜索引擎
[更多]

最近注册的项目

(09/11) webgraph
(08/31) cs101 fall 2007
(08/24) calal
(07/30) FanKa's Codelab
(07/02) seplatform
(07/02) CMajur
(06/28) vsta
(06/27) gossip
(06/27) cds_gossip
(06/21) mini search engine

查找: dyf



哈工大(威海) 新闻

发挥优势 突出特色 弥补不足 重在建设

新闻首页-头条新闻-学校要闻-视频新闻-图片新闻-高教动态-威海视听-党代会-学校首页-缅怀马祖光院士

当前位置: 新闻首页 >> 学校要闻 >> 科技信息 >> 新闻内容

微软亚洲研究院互联网创新暑期研讨会闭幕

【作者: 黄刚 来源: 计算机学院 浏览: 481 录入时间: 2007年8月25日】

微软亚洲研究院互联网创新暑期研讨会经过3天精彩的报告之后,于8月24日18点在N楼会议报告厅圆满结束。

在8月22日至24日,来自微软亚洲研究院的洪小文博士、Bill Dolan 博士、李世鹏博士、高剑峰博士、Hisami Suzuki 博士、Timothy Chiklovski 博士、张辉博士、陈正博士、张高博士、黄昌宁教授、Dean Slawson 博士、周明博士、牛成博士等专家分别就互联网创新与商业机会、Paraphrase and Textual Entailment、Media 2.0 - the New Media Revolution?、分布式系统的研发和展望、在线广告中的用户行为建模、中文分词及其应用、Web 2.0 时代的自然语言处理等方向作了报告,其中张高博士还着重在座的学生介绍了微软亚洲研究院“明日之星”的实习计划。专家们的精彩演讲和观众的积极提问,不断将会议推向一个又一个的高潮。报告之后,专家们还热情地为同学们解答问题。专家们旁证博引、热情洋溢的报告,让学生感到不虚此行。

在8月24日下午举行的研讨会上,还增加了一项题为“假如我有1000万”的有关学生自主创业的自由论坛,来自清华、北大、武汉大学、天津大学、哈尔滨工业大学等学校的8个团队参加了此次论坛,他们的想法新颖,内容丰富,涉及自主教学交流平台、多语言移动信息系统、对比评价导购网、信息融合与速查器、智能手持设备的信息交流应用、短片向FLASH的自动转换、video tour、U创等八项。最后经过专家组的评审,来自北京大学团队的“短片向FLASH的自动转换”项目获得第一名,清华大学的“video tour”团队和哈工大的“U创”团队分别获得第二名和第三名。

24日18点,林钦佑博士在N楼会议报告厅主持闭幕式,计算机学院常务副院长迟乐军对本次会议进行了总结,并衷心希望今后各位专家、教授以及学生能够来威海合作交流。最后林钦佑博士宣告本次会议圆满结束。

会议期间,校领导唐安阳、高学敏、马家辰分别同微软专家进行了会谈,并就后期的合作事宜进行了磋商,计算机学院、软件学院的有关领导参加了会谈。威海市高新区管委会副主任王政玉、刘勋显等会见了微软专家,就双方的进一步合作洽谈。

☆ 热点新闻 ☆

- ▶ 哈工大(威海)十大[2004]
- ▶ 威海校区参加校团委[2825]
- ▶ 一道谢维儒在我校[5030]
- ▶ 哈工大(威海)隆重[5025]
- ▶ 我校新增10名博导[4408]
- ▶ 谷桂林教授被聘为我[4098]
- ▶ 凸显“工程”特色[4074]
- ▶ 第十九届田径运动会[3988]
- ▶ 确保饭菜质量 稳定[3762]
- ▶ 毕业典礼在基[3615]
- ▶ 我校2007届研究[3604]
- ▶ 第十九届田径运动会[3380]
- ▶ 陈贺梅当选我校第四[3384]
- ▶ 2007届毕业生就[3197]
- ▶ 哈工大(威海)教授[2993]
- ▶ 2007年业委会[2928]
- ▶ 李绍流一行赴中国组[2910]
- ▶ 哈工大为昌平打[2844]
- ▶ 奖励学金颁奖大会[2818]
- ▶ 我校召开会议通报[2804]