

# 全唐诗分析系统 全宋诗分析系统

## 产品说明

中国国际图书贸易集团有限公司

北京大学数据分析研究中心

北京欣诺格科技有限公司

2010年8月

# 目 录

一、产品设计理念和定位.....	- 3 -
1、产品设计理念.....	- 3 -
2、产品定位.....	- 4 -
3、应用领域.....	- 4 -
4、典型用户.....	- 4 -
二、产品功能介绍.....	- 5 -
1、技术创新点.....	- 5 -
2、功能特点.....	- 7 -
三、系统配置.....	- 9 -
1、服务器.....	- 9 -
2、客户机.....	- 9 -
3、浏览器.....	- 9 -
四、价格体系.....	- 10 -
五、服务体系.....	- 10 -
六、联系方式.....	- 11 -

# 一、产品设计理念和定位

## 1、产品设计理念

“全唐诗分析系统”与“全宋诗分析系统”由北京大学数据分析研究中心和北京欣诺格科技有限公司联合研制。

1998年北京中文系承担了国家“九五”211项目——“全唐诗电子检索系统”的研发。项目主持人李铎博士为项目提出了三个目标：一、使用Unicode编码，使全球任何语言版本的操作系统均可正常使用该系统；

二、使用XML标记语言，脱离数据库平台，使用最先进的搜索引擎，实现跨平台的多维度检索；三、建设开放的、可扩展的模块化结构，可以任意增减数据。鉴于当初互联网上数据不精，往往给学术研究带来不便的问题，该项目组制定出文本校对高于纸介出版物的标准。此后，作为211项目子课题，吸收借鉴“全唐诗分析系统”的成功经验和先进的设计理念，又设计制作了“全宋诗分析系统”。



图：全唐诗分析系统



图：全宋诗分析系统

## 2、产品定位

“全唐诗分析系统”与“全宋诗分析系统”突破了以往全文检索的信息提供模式，在数据深层挖掘和知识发现方面具有开创性意义。其重出诗提取、格律诗标注、字及词组的频率分布统计、用户自作诗的格律分析等带有智能化特点。该系统还提供多维的检索分析方式，为中国古代文学、古代汉语、古文献学等研究领域提供了可靠的分析数据。它的成功研制标志着计算机科学在中文信息处理应用方面由全文检索的信息提供模式开始转向智能分析模式。“全唐诗分析系”与“全宋诗分析系统”是信息处理技术与中国古代文学研究的成功结合，展示出信息技术在中国古代文学研究领域应用的方向和广阔前景。海内外专家一致认为该系统达到了国际领先水平，在信息处理技术与中国古代文学研究结合方面居国际领先地位。

## 3、应用领域

- ◇ 在古代文学研究方面可应用于：诗人风格研究、诗歌流变研究、诗文相关研究、地域文化与诗歌特征研究等。
- ◇ 在古代汉语研究方面可应用于：古代诗歌语言模式分析、诗歌字及词组频率分析、诗歌语言发展史研究、音韵学相关研究等。
- ◇ 在古代文献整理方面可应用于：重出诗整理分析、误收诗分析等。
- ◇ 在古代文学教学方面可应用于：格律诗教学模拟系统研究、诗作评价、诗作相似性分析等。
- ◇ 在跨学科研究方面的可应用于：计算机技术与古代文学相关研究、古代文献数据深层挖掘、古代文献数据自动化整理等。

## 4、典型用户

目前全球范围内有近 100 家单位和个人正在使用或已经购买“全唐诗分析系统”与“全宋诗分析系统”，一些典型用户如下。

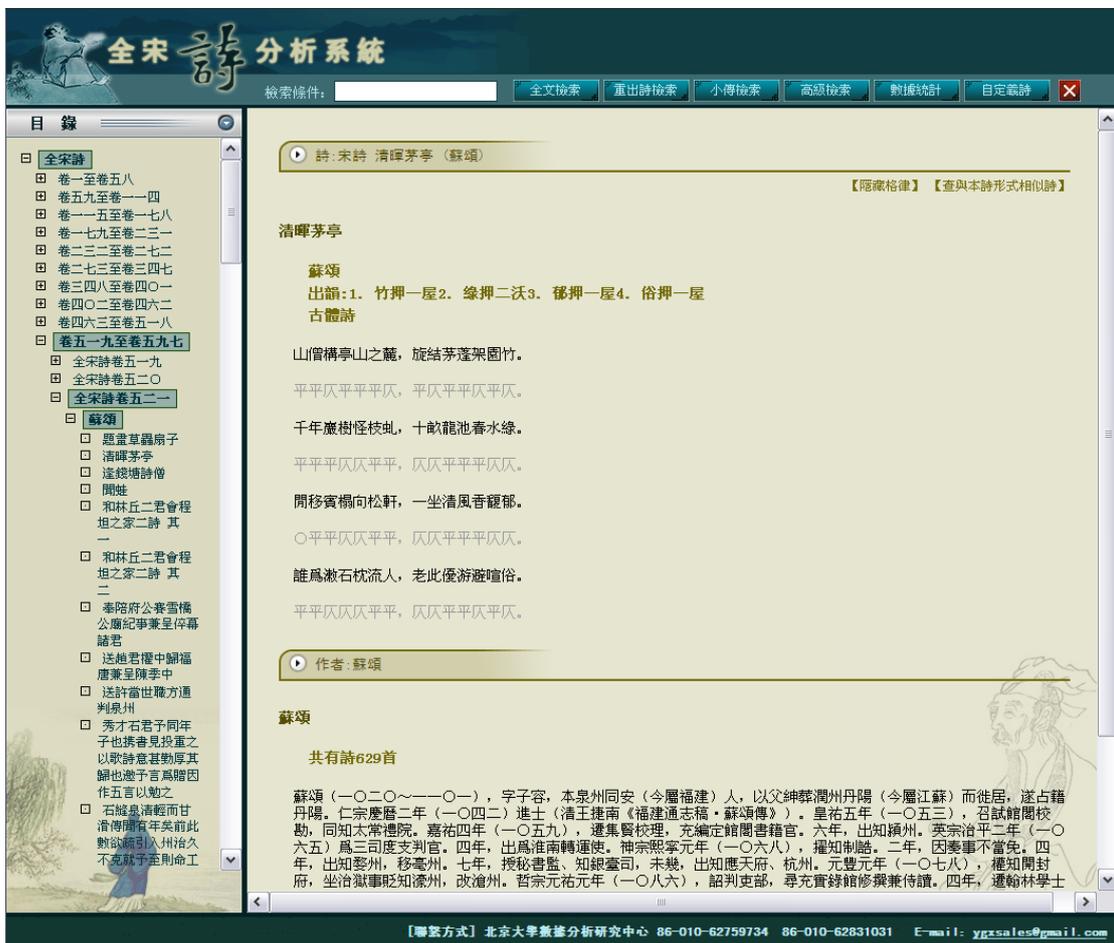
- ◆ 中国国家图书馆、中华书局、温岭市图书馆、台湾中央研究院、日本首都大学东京中文研究室
- ◆ 北京大学、浙江工业大学、西北大学、暨南大学、北京师范大学、北京语言大学、杭州师范学院、哈佛燕京学社、香港中文大学、高雄师范大学、台湾元智大学
- ◆ 北京珍本书店、北京科普图书有限公司

## 二、产品功能介绍

### 1、技术创新点

#### ◇ 格律诗自动标注

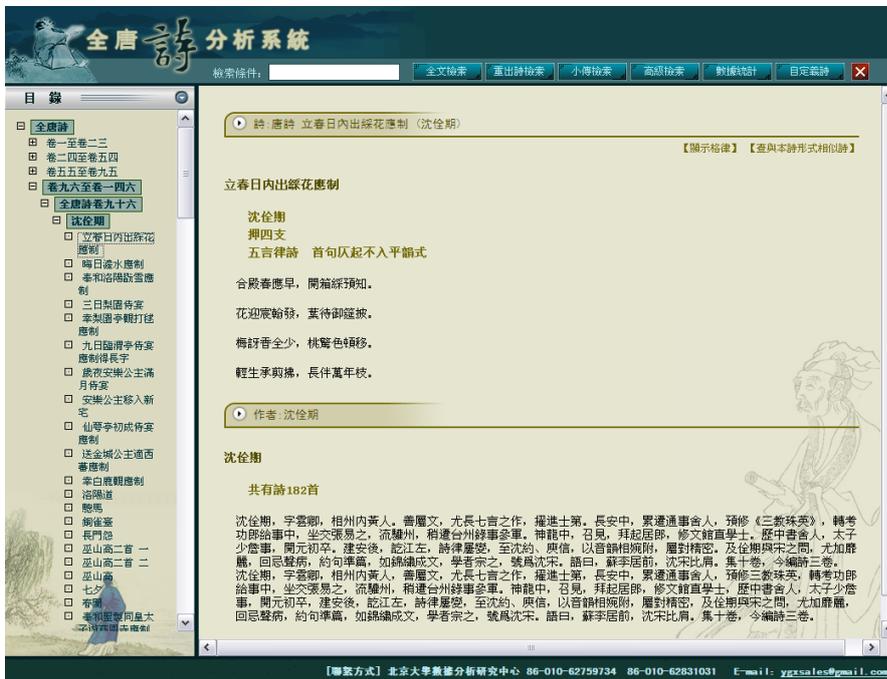
《全唐诗》57000 多首，《全宋诗》254240 首，其中格律诗占三分之二，数据量大，人工无法完成。本系统使用计算机自动完成此项工作，为此建设了格律诗模型库、音韵库等知识库，对《全唐诗》、《全宋诗》每一首诗进行多维判断，准确地提取出了全部格律诗。



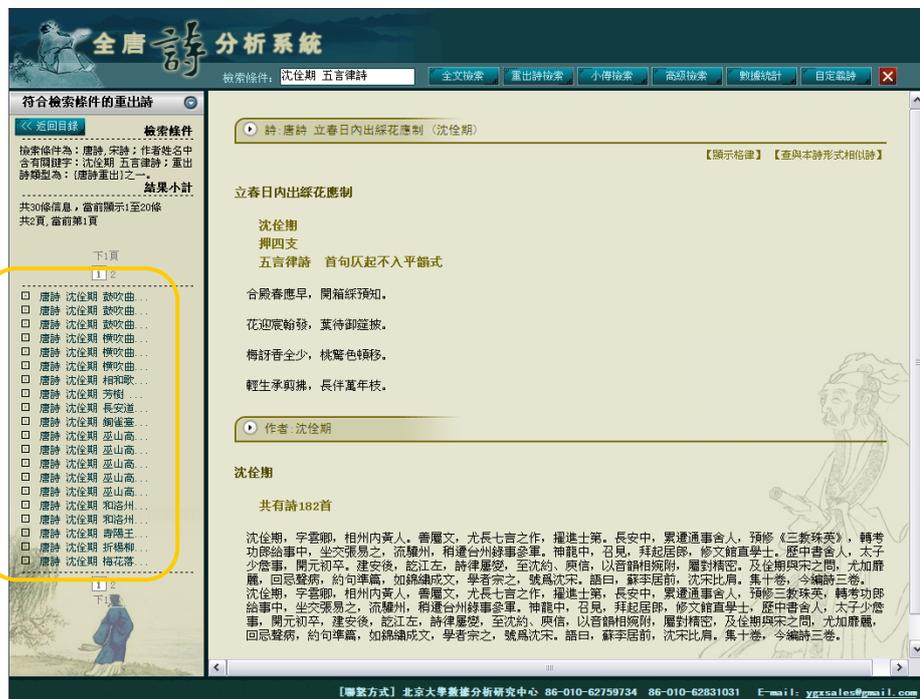
图：格律标注

## ◇ 重出诗的提取

《全唐诗》、《全宋诗》的编纂中有重出现象，本系统使用比较复杂的算法，由计算机自动提取出全部重出诗，为进一步整理分析《全唐诗》、《全宋诗》打下了基础。



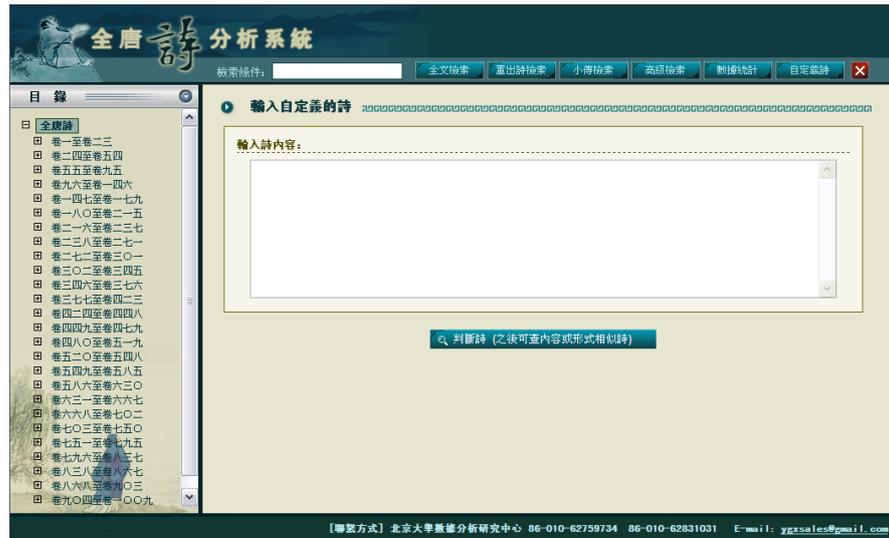
以“沈佺期 五言律詩”为检索条件检索重出诗



检索结果有 30 条信息

## ◇ 自作诗分析

本系统使用格律诗模型库、音韵库等知识库，对用户自作的诗进行分析，可以准确地分析出用户自己的诗作是否合律。



## 2、功能特点

### ◇ 强大的检索功能

“全唐诗分析系统”与“全宋诗分析系统”提供了全文检索、重出诗检索、诗人小传检索和高级检索。支持严格数据检索和混合模式检索，所谓混合模式是指输入繁体或者简体汉字，均可检索到与之相对应的简体或繁体字，如输入“后”可以检索到“后”与“後”。所谓严格数据检索，即输入“后”只检索“后”，输入“後”只检索“後”。该系统支持人名、别名扩展检索，如输入“杜甫”或“子美”或“杜工部”均可检索到全部杜甫诗作。



## ◇ 科学的分析功能

“全唐诗分析系统”与“全宋诗分析系统”统可以进行字、词组和诗歌格律使用情况的统计和分析。作为智能分析系统，它远远超过了人力统计分析的效率，利用计算机进行智能分析和穷尽式统计，可以为学术研究提供更多的增长点。



## 三、系统配置

### 1、服务器

中央处理器：Intel Celeron Dual-Core 2.4GHz（或以上）

内存：2GB DDR3（或以上）

硬盘：200GB（或以上）

网络带宽：1Mbps（或以上）

操作系统：Windows 2003（或以上）

Java 运行环境：JDK1.6（或以上）

数据库运行环境：MySQL 5.0（或以上）

### 2、客户机

中央处理器：Intel Celeron 2GHz（或以上）

内存：512MB DDR2（或以上）

网络带宽：512Kbps（或以上）

显示器分辨率：1024×768（或以上）

操作系统：Windows XP（或以上）/ Windows 2003（或以上）、

浏览器：IE 6.0（或以上）/ FireFox 3.5（或以上），须开启 Javascript 功能

字体库：Sun-ExtA、Sun-ExtB、宋体-方正超大字符集

多媒体插件：Adobe Flash Player V9.0（或以上）

### 3、浏览器

IE 6.0（或以上）/ FireFox 3.5（或以上），须开启 Javascript 功能

字体库：Sun-ExtA、Sun-ExtB、宋体-方正超大字符集

多媒体插件：Adobe Flash Player V9.0（或以上）

## 四、价格体系

产品类型	税前年服务价格 (元人民币)	IP 段 类型	IP 段 数量	用户 数	配套费用
服务型产品	¥2,000.00	C	1	1	每增加一个用户，年服务费增加贰仟元人民币（¥2000.00）

## 五、服务体系

### 技术服务内容:

- 1、工作时间人工在线服务支持（Tel: 86-010-82773903 Fax: 86-010-82790956 QQ: 1533036511）。
- 2、365\*7\*24 离线服务支持（Email: xng\_support@163.com MSN:xng\_support@hotmail）。
- 3、24 小时内排除一般故障。
- 4、重大事故期间为用户提供备份网络服务。
- 5、1 个工作日内答复用户投诉。

## 六、联系方式

地 址：图书馆服务中心·中国国际图书贸易集团有限公司  
海淀区车公庄西路 35 号  
中国·北京 100048

电 话：010-68433074

传 真：010-68412023

电子邮件：tsg@mail.cibtc.com.cn

网 址：<http://www.cibtc.com.cn/>