

乐思信息采集系统

针对任意网页自动采集精确内容
全球海量信息资源为您随心所用



深圳市乐思软件技术有限公司

<http://www.knowledsys.com>

网络淘金

- **Google**采集全球因特网网页数据库，成就800亿美金市值的奇迹
- 你想从因特网的公开信息中采集以下数据库从而获取企业情报与提高销售业绩吗？
 - 企业名录，产品资料...
 - 行业新闻，供求信息...
 - 实时行情，论坛帖子...
 - 网络博客，科研数据...

关键问题

- 大量数据分散地在各个网站中以不规则形式显示
- 而你需要的不是纷繁复杂的网页形式，而是一个集成的纯净的精确的数据库
- 以此支持你的网站，营销，研究，数据挖掘，CRM，ERP，竞争情报系统



能从这信息海洋中
淘出金沙，铸成金子吗？



乐思信息采集系统

让梦想成为可能！

工作原理



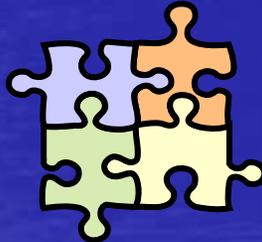
从大量网页
批量采集数据
到数据库中

Google做不到的，我们可以！

- Google采集的仅是因特网海洋中**表层的网页**
- 乐思信息采集系统却可以做到Google做不到的，采集隐藏在登录框与查询表单后面的**深网网页中的精确字段内容**，保存记录到**数据库中**

强大的定制能力

- 可针对任意网页采集，而非事先设定，让你的数据源源不绝，随时更新
- 针对任意数据格式，预设多种抽取方法，供你选用

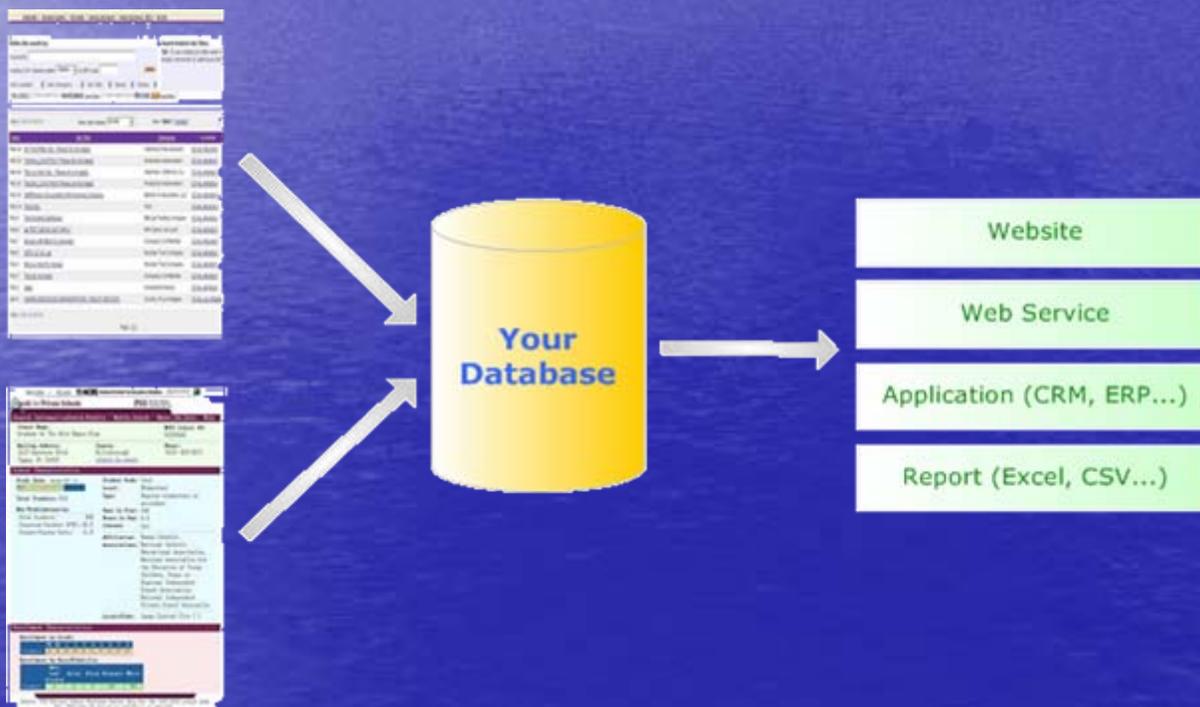


主要功能

- 根据用户自定义的任务配置，批量而精确地抽取因特网目标网页中半结构化与非结构化数据，转化为为结构化的记录，保存在本地数据库中，用于内部使用或外网发布。



从网页采集整合精确内容以便利用



信息采集 快、准、狠

- 快：每小时可采集几万到几百万条记录



- 准：字段级精确抽取与处理



- 狠：采集批量化与自动化



系统特点 (1)

- 采集方法的灵活性与采集数据的准确性
灵活性: 可处理任意复杂查询与页面
准确性: 结果数据高度准确(99%-100%)
- 对目标网站进行信息自动抓取, 支持HTML 页面内各种数据的采集, 如文本信息, URL, 数字, 日期, 图片等
- 用户对每类信息自定义来源与分类
- 可以下载图片, PDF与其他各类文件



系统特点（2）

- 支持下一页自动浏览功能
- 支持表单查询
- 支持动作脚本
- 支持从一个页面中抽取多个数据表
- 支持数据的多种后期处理方式



系统特点（3）

- 支持智能替换功能，可以将内容中嵌入的所有无关部分如广告去除
- 支持多页面文章内容自动抽取与合并
- 支持数据库表结构完全自定义，充分利用现有系统
- 支持多个栏目的信息采集可用同一配置一对多处理
- 保证信息的完整性与准确性，绝不会乱码



系统特点（4）

- 支持用户名与密码自动登录
- 支持记录唯一索引，避免相同信息重复入库
- 数据直接进入数据库而不是文件中，因此与利用这些数据的网站程序或者桌面程序之间没有任何耦合
- 附带丰富的采集示例，便于快速学习与使用
- 支持命令行格式，可以Windows 任务计划器配合，定期抽取目标网站



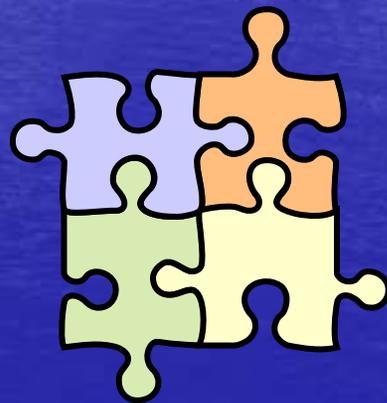
运行环境

- 操作系统：Windows XP/NT/2000/2003
- 内存：最低64M内存，建议256M或以上
- 硬盘：最少20M空余硬盘空间



乐思信息采集系统

助您整合全球海量信息资源



深圳市乐思软件技术有限公司

关于乐思软件

- 深圳市乐思软件技术有限公司(简称乐思软件)成立于2003年5月, 专业从事网络信息采集服务与产品开发, 在网络信息采集领域拥有丰富的实践经验与国际领先的技术优势
- 欲了解关于乐思软件的更多信息, 请访问我们的网站: <http://www.knowlesys.com>

联系方式

- 公司名称：深圳市乐思软件技术有限公司
- 联系人：徐先生

- 电话：0755-86032826
- 电邮：xuyong@knowlesys.com
- 网站：<http://www.knowlesys.com>

- 技术支持：电话，MSN，Skype，电子邮件