

# 德国*e-Humanities*项目*TextGrid* 的概述与启示

曾勤

北京大学图书馆

# 目录

- ◎ 网格(Grid)
- ◎ TextGrid概述
- ◎ TextGrid具体应用案例
- ◎ 对中国e-Humanities的启示

- ◎ 科学技术的进步、高性能计算的需求、空闲资源的利用、计算结果的共享等导致了计算机技术领域的又一次重大突破--网格（Grid）技术的发展。
- ◎ 网格技术：实现各种资源的全面共享，消除资源“孤岛”，为用户提供一体化信息和计算、存储、访问等应用服务。

## 代表性的全球网格项目：

- ◎ 亚太地区的ApGrid(Asia Pacific Grid)
- ◎ 美国的TeraGrid
- ◎ 欧盟的DATAGrid和EuroGrid
- ◎ 英国E-Science计划
- ◎ 德国Deutschland Grid
- ◎ 中国国家网格 (CNGrid)
- ◎ 中国教育科研网格 (ChinaGrid)

主要构建重要科学领域的全球协作基础设施，  
应用于生物、医学、物理、地理等自然科学  
领域。

# *e-Science*和*e-Humanities*

- ◎ e-Science: 科研人员组成跨学科、跨组织、跨地域的虚拟组织, 开展科学合作和创新研究。
- ◎ e-Humanities: 等同于人文领域的e-Science。提供一个虚拟的网络和跨学科环境, 通过分布式资源、协同合作的方法和标准化的网格计算, 进行学术文本数据处理、分析、注释、编辑和链接。

# *TextGrid*概述

- ◎ D-Grid: 即Deutschland Grid, 将网格计算应用于科学研究, 包括气象、高能物理、生物医学研究、生命科学应用和复杂工程解决方案等领域。
- ◎ TextGrid: D-Grid计划框架下唯一应用于人文科学领域的网格项目, 旨在创建一个虚拟研究图书馆, 为人文学者提供处理、分析、注释、编辑、链接和发布文本数据的网格工作平台。

# 为人文学科研究提供以下人文基础设施

- ◎ 学术文本处理领域，运用实证方法（empirical methods）创建和使用数字文本，解决传统的和新出现的问题
- ◎ 利用分布式资源，引进合作方法和标准化工具提高文本数据处理水平
- ◎ 提供易于安装和使用的图形用户界面（graphical user interface，简称GUI）和开放式接口、统一出版平台、允许用户定义工作流程和数据结构
- ◎ 自动完成如符号标记，创建索引，词语分类，结构分析，类型设置等一系列步骤
- ◎ 即使相距遥远、操作系统或编程语言不同，能够提供一个统一的平台，允许用户对自己的模块进行扩展，对不同模块和数据整合和无缝链接
- ◎ 提供访问资源的透明度和安全性，为研究者创建开放式的科学研究、大范围的资源共享和广泛的异地协同合作的服务平台

# *TextGrid*具体应用案例

## 文献编辑

- ◎ 支持不同阶段各个版本的创建
- ◎ 对历史性的手稿和出版物数字化，并链接到存储网格，进行存储和访问
- ◎ 对原始资料图片和可处理的文本进行链接，帮助用户阅读手稿图片
- ◎ 提供从可检索和已注释的文本到扫描手稿相应部分的导航
- ◎ 通过部分自动操作（如分割原始图像）和整合链接过程，简化繁琐的编码任务
- ◎ 利用网格计算，自动整理和实时更新作品起源及评论等信息
- ◎ 支持共同创建注释和统一的书目信息，并应用于全面检索、生成网页和自动翻译等.....

# *TextGrid*具体应用案例

## 词典编纂学和语言学的应用

- ◎ 利用计算语言学、信息检索和网格计算，提供一个支持德国历史语言学和现代德语文集编辑的综合平台
- ◎ 建立展现和管理不同地区和时代的词汇数据的词典编纂者平台
- ◎ 通过智能服务对TextGrid资源提供增强型访问来满足不同语言学学科的需求：
  - 语源学：搜索外来词汇
  - 方言学：搜索区域异体字
  - 篇章语言学：搜索特定的文本类型(如诗歌)

- ④ 发展了不同时期的词形学分析工具，自动发现不同词形的词位
- ④ 通过词条、部分语句、地区和语言的形态句法信息丰富每个文集的标号内容
- ④ 依据对称和转换，扩大词汇入构项之间由于原始的交叉参考词汇所产生的词群
- ④ 通过标准界面提供统一的搜索功能，如分类搜索中心词，语法信息以及对不加限定的纯文本、基于原有和新插入参考文献的语义上相关术语的搜索，大大提高了字典的实用性

# 对中国*e-Humanities*的启示

- ◎ 中国教育科研网格（ChinaGrid）：整合了全国二十多所高校的高性能计算资源、存储资源、数据资源、知识资源等，通过有序管理和协同计算，消除信息孤岛，发挥综合效能，实现资源的广泛共享、有效聚合、充分释放，为前沿科学研究探索提供了必备手段和条件。
- ◎ 借鉴德国D-Grid的发展计划与应用，是否可以建立ChinaGrid下的“ChinaTextGrid”，为人文学科研究提供*e-Humanities*的基础设施，对丰富的人文资源进行开放式的科学研究、大范围的资源共享和广泛的异地协同合作。同时也是对数字图书馆建设应用网格技术提供一体化的平台、实现资源的全面共享、解决海量数据处理、进行资源集成和知识管理的一个新的挑战。

D-Grid: <http://www.d-grid.de>

TextGrid: <http://www.textgrid.de>

谢谢！

Thanks,  
for your  
attention !

