

973课题“物理与网络空间的协同感知” 数据平台

中科院自动化研究所

2012年9月3日



提纲

- 数据平台简介
- 茉莉花革命数据



数据平台简介



数据平台访问

- 自动化所内:

 - IP:192.168.73.99

- 自动化所外:

 - IP:159.226.19.39

- 数据库: MySQL (GUI/命令行)

- 分配访问账号

- 可联系:

wqmin@nlpr.ia.ac.cn; jtsang@nlpr.ia.ac.cn.



数据格式

- 数据库：通过某种实体组织**关联**不同源数据
(**主题/地点/群体/时间...**)；
- 表：每种源数据构成一个表；
- 记录：每个条目写成记录
(**图片、视频、Tweet、新闻...**)；
- 主键名：实体_数据源_记录号



茉莉花革命数据



数据概况

- 以茉莉花革命为实体索引组织Flickr、Youtube、Twitter、Google News数据；
- 七个国家、105个事件；

| 数据源 | #记录 | 主要属性 |
|-------------|--------|---------------------------|
| Flickr | 42,704 | 标题、描述、评论、标注、作者、时间、Geo_tag |
| Youtube | 1,434 | 标题、描述、标注、作者、时间 |
| Twitter | 4,217 | 时间、点击数、转发数、评论、作者 |
| Google News | 686 | 发布源、时间、图片/视频ID |

研究计划及存在问题

■ 研究计划：

- 主题发现和数据关联；
- 个人和群体挖掘分析；

■ 存在问题：

- 只有正样本；
- 物理空间信息不足。



谢谢！

