

Web Information Acquisition, Data Integration and Its Application in Crisis Management



 **Jiping Liu**

Chinese Academy of Surveying and Mapping



Outline

01

Background

02

Current Situation and Problems

03

Research Methods

04

Typical experiments

05

Conclusion and Ongoing work

☞ Background

Disaster: affect the world every day

- A grave occurrence having ruinous results. (Webster's Dictionary)
- It could be **natural or man-made**.
- such as flood, hurricane, fire, earthquake, plane crash, etc.

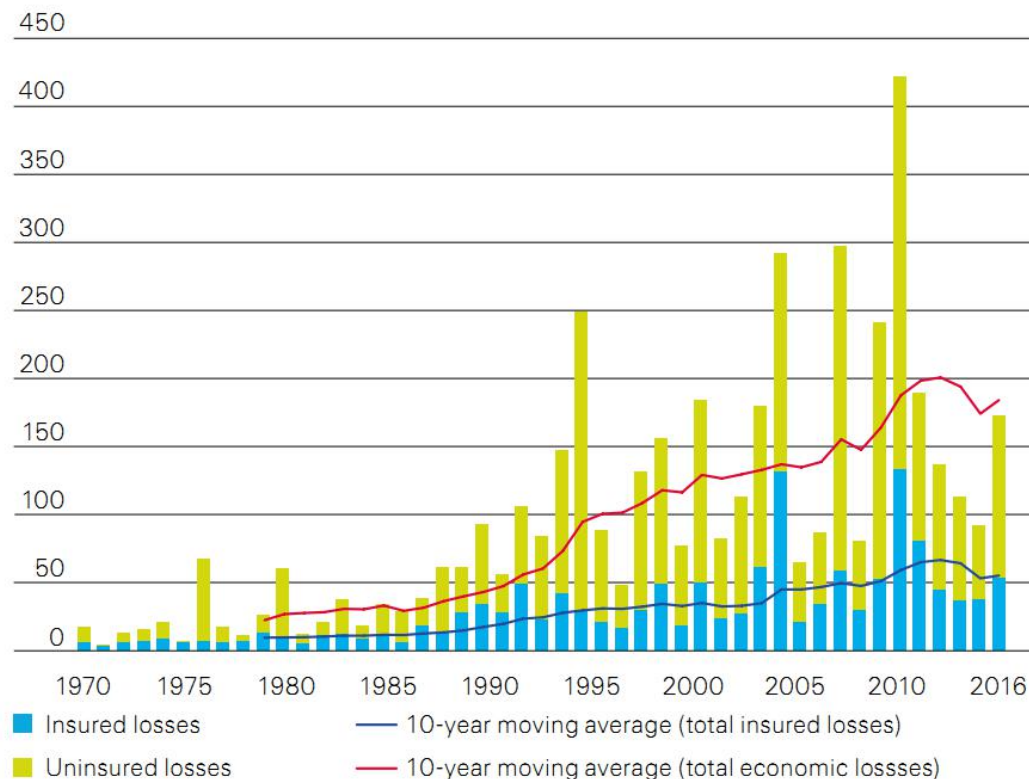


➡ Background

- Swiss Re's sigma study reveals the number of catastrophic events and victims and the economic losses from 1970 to 2016:

The estimated total economic losses were USD 175 billion in 2016. Catastrophe losses in 2016 were 0.24% of global GDP, again in line with the 10-year average.

Economic losses, 1970–2016, USD billion in 2016 prices



Total losses = insured + uninsured losses.

Background



Terrorist attacks



Flood



Tsunami



Hurricane



Earthquake

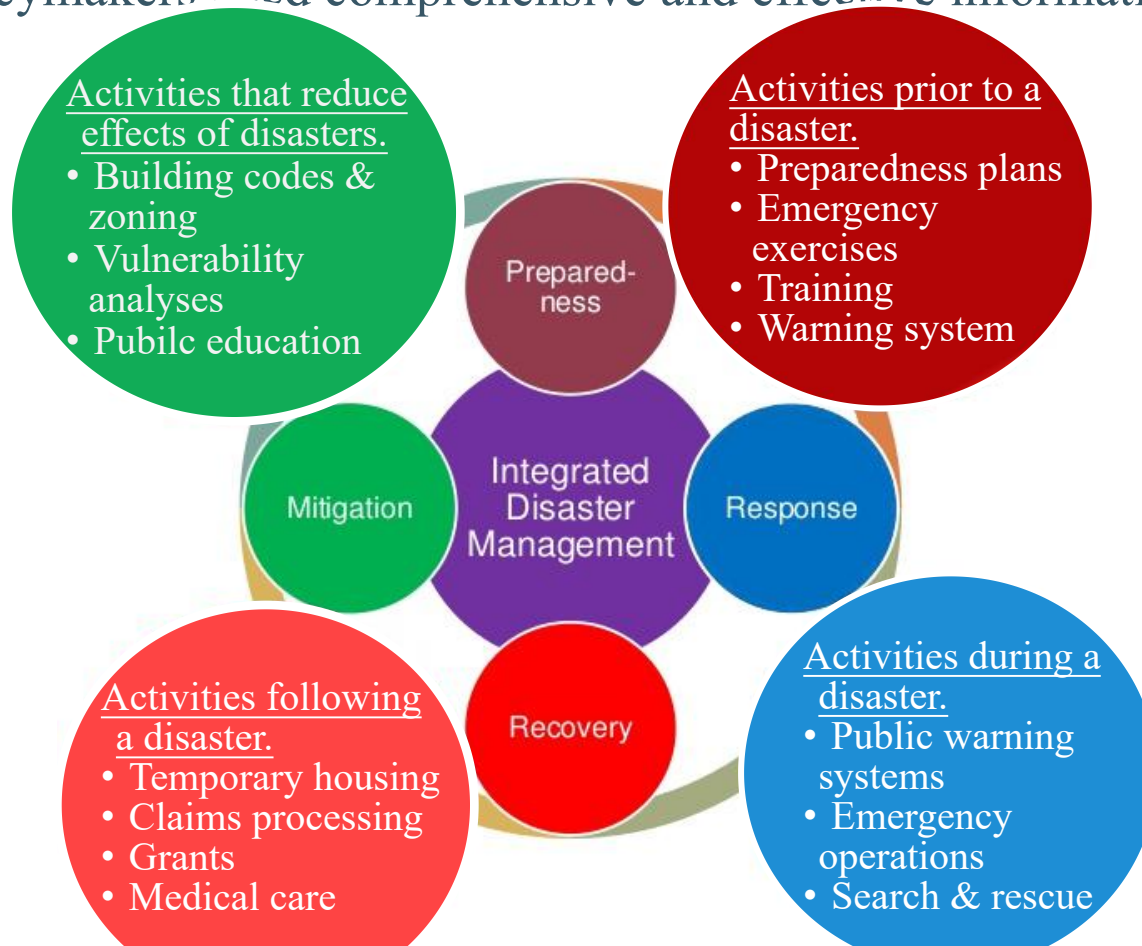
In china, only the natural disasters caused 1432 people died and direct economic losses stood at 503.29 billion yuan in 2016.

Background

How to respond better to the disaster?

How to deal with emergency in a short time?

Policymakers need comprehensive and effective information to make decisions



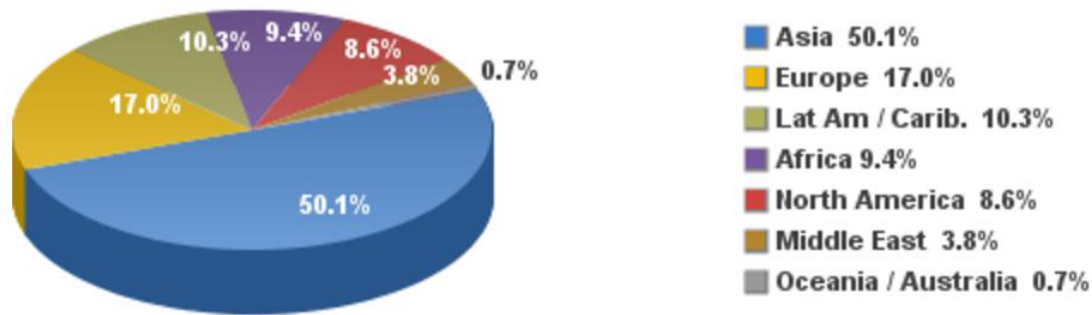
A continuous process of planning and coordinating for disaster management.

👉 Background

Internet of Things: The new era is coming

- Up to March 31, 2017, the number of internet users worldwide was 3.7 billion, up from 3.4 billion in the previous year.

**Internet Users in the World
by Regions - 2017 Q1**



- The Asian region ranked first, which was 50.1% of the total
- The number of netizens in China has reached 731 million. Penetration is reached 52.7%

Background

- As the important channels of information release, dissemination and exchange, internet contains a large number of rich geographical information.

The collage consists of three main screenshots:

- Top Left:** A Google search for 'restaurants' in Beijing. The results list several restaurants with ratings and addresses, such as 'Da Dong Roast Duck Restaurant' and 'Li Qun Roast Duck Restaurant'. A map shows the location of these restaurants in the Dali Courtyard area.
- Top Right:** A 'street view' image of Tiananmen Square, showing the Tiananmen Gate and the flagpole. A speech bubble labeled 'street view' points to the image.
- Bottom:** A 'Beijing Real-Time Traffic Information' map. The map shows the city's road network with color-coded lines indicating traffic conditions: red for congestion, yellow for moderate traffic, and green for clear roads. A speech bubble labeled 'Beijing Real-Time Traffic Information' points to the map.

Overlaid text boxes include:

- restaurants** (pointing to the Google search results)
- street view** (pointing to the Tiananmen Square image)
- Geographical information on the internet is increasingly abundant** (centered over the traffic map)
- Beijing Real-Time Traffic Information** (pointing to the traffic map)

At the bottom of the traffic map, there is a legend for '实时路况' (Real-time traffic conditions) with categories: '更改时间' (Change time), '拥堵' (Congestion), and '畅通' (Clear). A scale bar indicates '5 英里' (5 miles).

Background

As a kind of geographic information public service, Large volume of Point of Interest (POI) are provided by many geographic information services on the Internet, and the number will increase rapidly in future.

- POI (Point of Interest) is mostly used in GIS and LBS
- Large volume of POI are provided by many web sites
- POIs are updated periodically

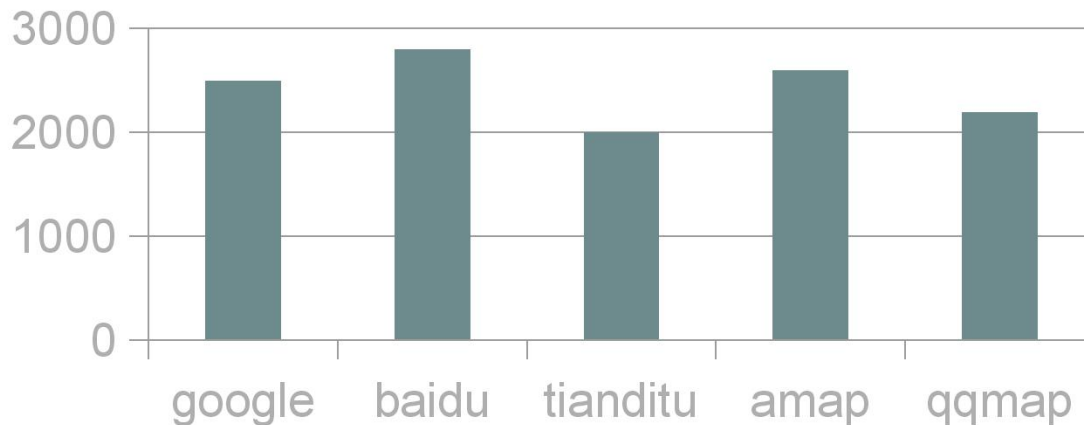


More than 1000 geographic information service websites

➡ Background

As a kind of geographic information public service, Large volume of Point of Interest (POI) are provided by many geographic information services on the Internet, and the number will increase rapidly in future.

- POI (Point of Interest) is mostly used in GIS and LBS
- **Large volume of POI are provided by many web sites**
- POIs are updated periodically



The average number of POI for each website is more than **20 million**, and the total of POIs from the five web sites is up to **121 million**.

Background

As a kind of geographic information public service, Large volume of Point of Interest (POI) are provided by many geographic information services on the Internet, and the number will increase rapidly in future.

- POI (Point of Interest) is mostly used in GIS and LBS
- Large volume of POI are provided by many web sites
- **POIs are updated periodically**

| Name | Frequency of updates |
|--------------|----------------------|
| Google Map | About once a month |
| Baidu Map | Once every quarter |
| Tianditu Map | Twice a year |
| AMap | Once every quarter |
| QQMap | Once every quarter |

Background

Sources of internet geographical information:

criteria

Data from
authoritative and
professional

Public service data

Data of GPS

Data
spontaneously

Data derived
from Web2.0

In

In

Di

Cr

Ir

Up

Up

netizens

Geotags

vague

lower

lower

weaker

no

crowdsourcing、
passive

crowdsourcing、
passive

crowdsourcing、
passive

2016年10月14日 星期五

中国地震信息网
China Seismic Information

地震资讯 地震背景 观测网络 观测技术 预报预报 地震科
数据共享 应急救援 地震文案 法规标准 市县工作 地震科

数据共享 全国2级以上地震(19)

全球7级以上地震
全球6级以上地震
全国5级以上地震
全国2级以上地震

日期: 2015/01/01 至 2016/1
纬度: -90.0 至 90.0
经度: -180.0 至 180.0
震级: 5 至 10

中国地震局
CHINA EARTHQUAKE ADMINISTRATION

首页 新闻资讯 政务公开 服务办事 交流互动 地震频道 资料库 搜索

Table & Text from authoritative and professional departments

东台县海域发生5.9级地震

发布时间: 2016-10-07 07:49:39 信息来源: 中国地震局办公室

据中国地震台网测定,北京时间2016年10月6日23时51分,在台湾台东县海域(北纬22.60度,东经121.38度)发生5.9级地震,震源深度约20公里。震中距离台湾最近海岸线约30公里,距最近岛屿绿岛约12公里,距台东县约40公里,距台东市约30公里。

行业动态
市县工作
媒体播报
热点报道
重要专题
震情速递

网站地图 | 网站声明 | 联系我们

中华人民共和国计算机信息网络国际联网单位备案号
京ICP备06029777号 中国地震局版权所有 建议使用IE浏览器浏览本网站

下表6、最新10-

| 发震时间 | 纬度(°) | 经度(°) | 深度(千米) |
|---------------------|-------|-------|--------|
| 2016/10/13 23:47:28 | 34.1 | 96.8 | 8 |
| 2016/10/13 22:52:16 | 24.6 | 98.9 | 6 |
| 2016/10/13 22:46:03 | 36.4 | 105.6 | 4 |
| 2016/10/13 22:14:60 | 23.8 | 98.8 | 6 |
| 2016/10/13 22:04:46 | 25.1 | 98.1 | 6 |
| 2016/10/13 21:52:36 | 27.0 | 103.4 | 6 |
| 2016/10/13 20:59:30 | 24.5 | 113.1 | 7 |
| 2016/10/13 20:16:07 | 37.0 | 92.1 | 0 |
| 2016/10/13 19:41:46 | 41.2 | 112.8 | 13 |
| 2016/10/13 19:39:27 | 37.3 | 111.8 | 14 |

下图7、最新10个地震震中分布图

Background

Sources of internet geographical information:

网易新闻 网易首页 - 新闻 - 体育 - 娱乐 - 财经 - 汽车 - 科技 - 数码 - 手机 - 女人 - 房产 - 游戏 - 读书 - 论坛 - 视频 - 博客 - 乐乎

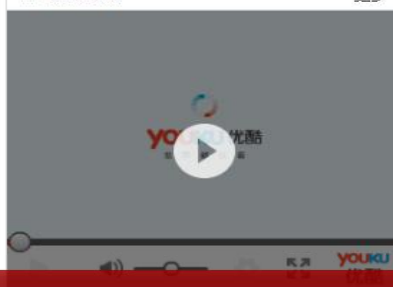
请输入关键词 新闻

四川雅安 地震伤亡地图 发生里氏7.0级

专题首页 | 最新消息 | 图片报道 | 视频报道 | 系列策划 | 部分遇难

视频播报

更多



芦山县遭遇缺水困境 8万村民瓶装水度日

道路抢通了，救援物资也进来了。现在，对震中的居民来说，最大的困难便是饮水问题。自来水管网基本破坏殆尽，井里河里虽然有水，因为担心水质水质问题也尽量避免饮用。...
[详细]

四川雅安地震最新消息

更多

- 地震专家称雅安地震5级以上余震还会继续发生
- 四川雅安7级地震震感强烈 居民称大里房屋坍塌
- 四川雅安7级地震 成都铁路局停运82列车
- 专家：四川雅安地震与汶川地震无关
- 邛崃市震雅安芦山县地震影响 2人死亡 30人受伤
- 四川雅安7级地震最新消息 44人死亡 500人以上受伤

雅安人民医院：0835-2862106
成都消防总队电话：0835-119
雅安救灾物资捐赠电话：0835-2242305/2362015

全网寻人
寻找每一个生命奇迹 或者家人

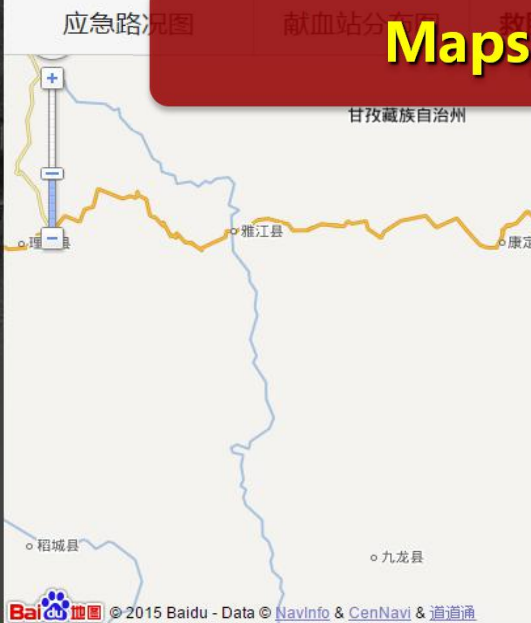
地震常识专题



四川雅安芦山县简介
芦山县（四川省雅安市所辖县）位于四川盆地西缘，属盆周山区。北与汶川，东北与崇州、大邑，东南与邛崃，南与雨城区，西南与天全，西北与宝兴相连。县城距雅安31公里...
[详细]

欲知更多地震常识，请进...

Maps from Internet service enterprises



灾区地图分布

详细

地震实景航拍图

物资需求分布

强度及人群分布

应急路况图

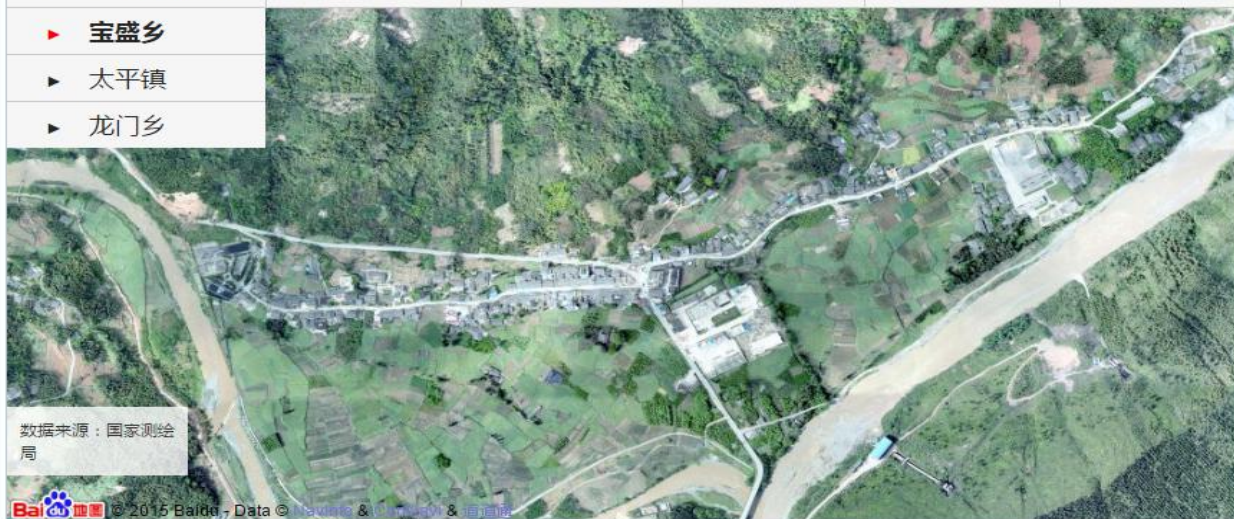
献血站分布图

救助站分布图

▶ 宝盛乡

▶ 太平镇

▶ 龙门乡



数据来源：国家测绘局

Baidu 地图 © 2015 Baidu - Data © NavInfo & CENavi & 道道通

Update mode

initiative、regular

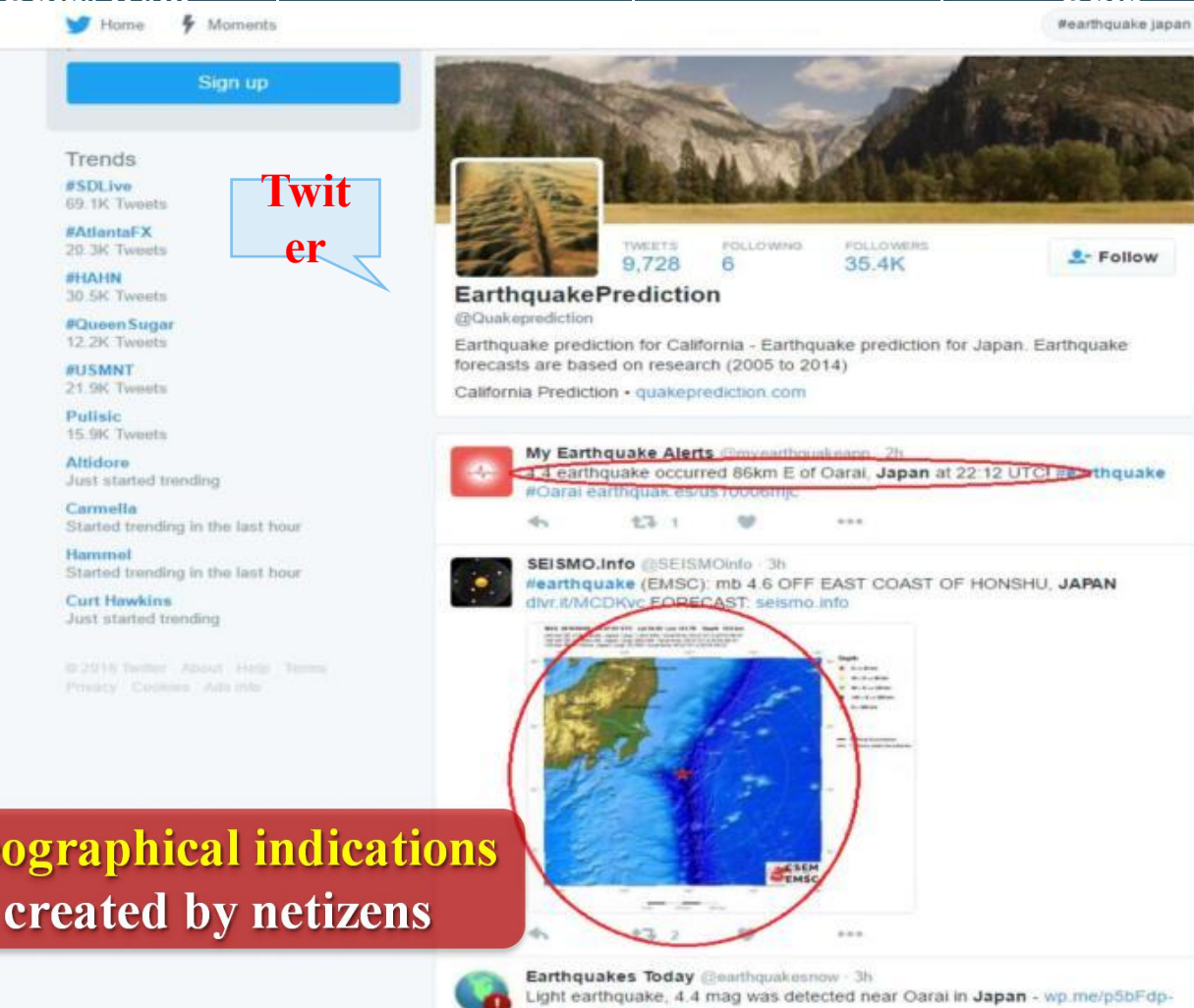
👉 Background

Sources of internet geographical information:

| criteria | datas from | Data | Data derived from Web2.0 |
|------------------------|------------|---|--------------------------|
| Information sources | spec | 系甘噶喇 | netizens |
| Information forms | | 我在这儿，我很好！ 我在： 大观北路 广州市天河区 | Geotags |
| Directivity of topic | | 1分钟前 来自:小米手机1S | vague |
| Credibility of content | | 信息时报 | lower |
| Geometric richness | | 【📢 呼吁：请关注灾区女同胞，女性用品匮乏】@新华社中国网事：抗震救灾中，女同胞还得面对“例假”的麻烦。卫生巾、止血药、生理盐水、流动厕所等这些满足女性特殊需求的用品也很重要 | lower |
| Integrity of attribute | | 五年前的汶川地震就女性用品的问题。希助时多关注灾区女性。小捐助， | weaker |
| Update subject | | geographic data from GPS receiver | no |
| Update mode | init | | owdsourcing、passive |

Background

Sources of internet geographical information:

| criteria | aut | Datas from | | Data | |
|------------------------|-----|---|--|------|------------------------|
| Information sources | spe |  | | | ta derived m Web2.0 |
| Information forms | | | | | netizens |
| Directivity of topic | | | | | Geotags |
| Credibility of content | | | | | vague |
| Geometric richness | | | | | lower |
| Integrity of attribute | | | | | lower |
| Update subject | | | | | weaker |
| Update mode | ini | | | | no |
| | | | | | /dsourcing、 passive |

**Geographical indications
created by netizens**

passive

passive

Background

Characteristics of internet geographical information:

Search ON

Front Page |

2004: The Asian tsunami

Massive sea earthquake have killed or have killed or southern Asia feared dead.

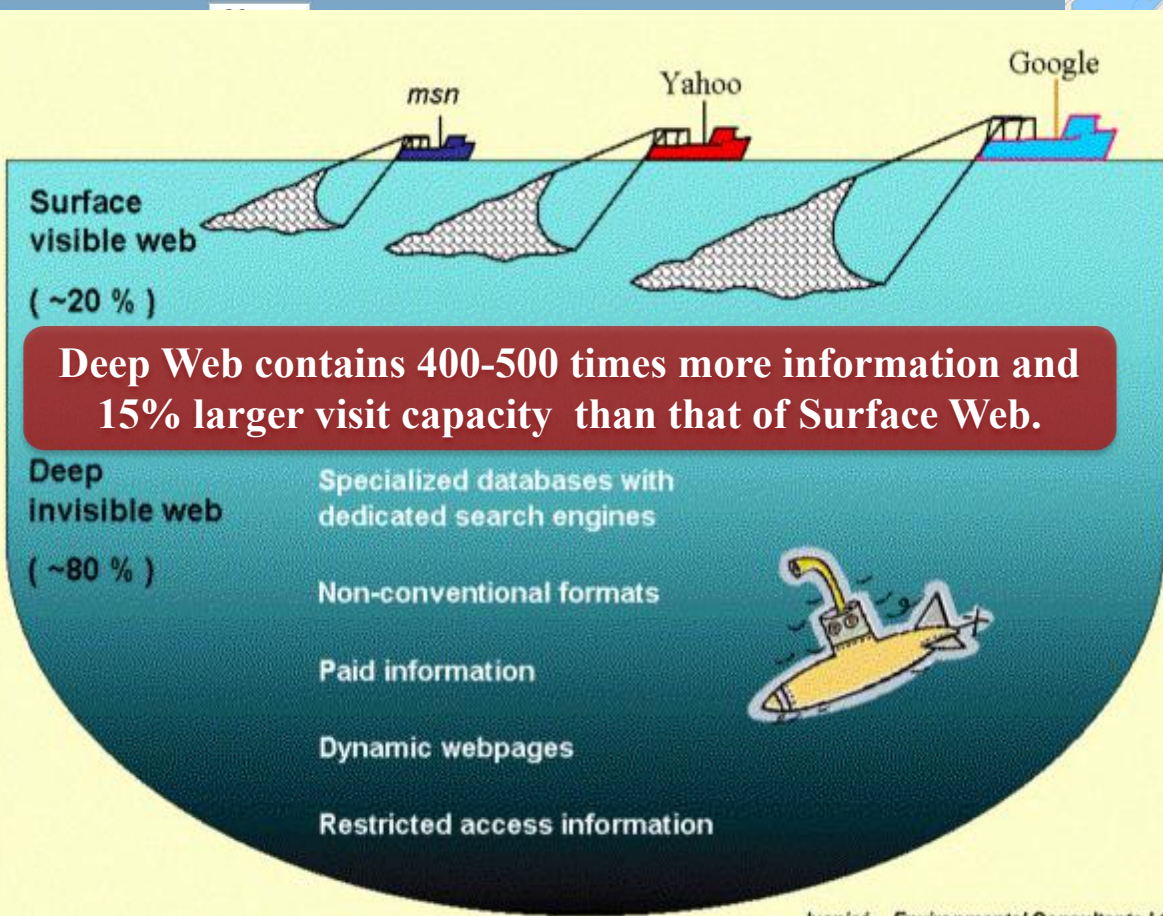
An 8.9 magnitude earthquake at 0759 local time generated the world has seen in years.

The wall of the Indian Ocean slammed into little or no warning.

Officials in India have already the thousand

expected to rise sharply over the next few days.

200,000 people in 13 countries. At least 128,000 people died in Indonesia alone.

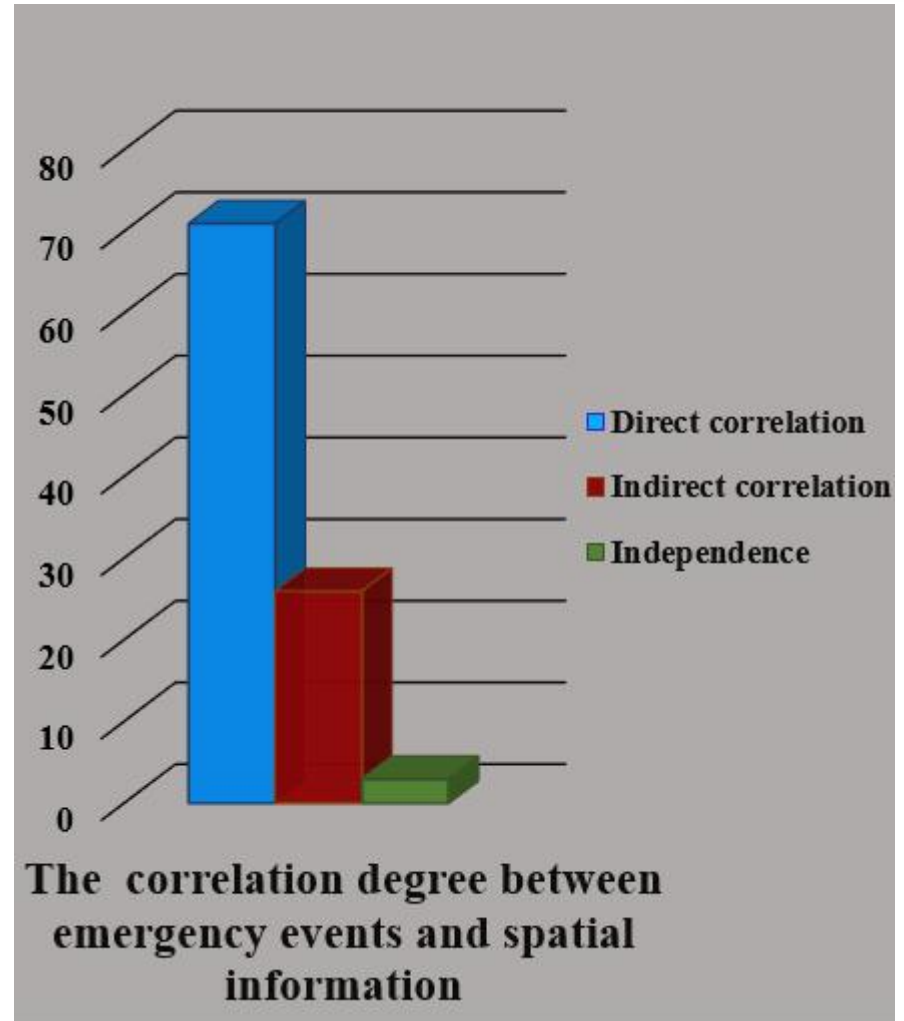


Juanicó – Environmental Consultants Ltd.



Background

The direct correlation degree between emergency events and spatial information is 71%, and the indirect correlation degree is 26%.



👉 Background

Geographical information on the Internet has played a very important role in all aspects of disaster response

➤ pre-disaster

- provide the early warning

➤ during disaster

- providing real-time information
- helping coordinate relief efforts
- finding missing persons

➤ post-disaster

- provides effective communication
- launching donation drives

An Ongoing Response to Hurricane Sandy

Recovering and Rebuilding after Hurricane Sandy.

The response to Hurricane Sandy began while the massive storm was still gaining strength in the Atlantic Ocean. By the time it made landfall along the East Coast, FEMA teams and resources were already in place to begin helping the millions of people that would be affected by the storm's immediate impact. Recovery and clean up efforts have only intensified since then, as federal agencies, local governments, first responders and volunteers work together to recover, rebuild and come back stronger than before.

The Hurricane Sandy Rebuilding Task Force's Rebuilding Strategy lays out a series of recommendations that will help the Sandy-impacted region rebuild in a way that will prepare them for these impacts – and that will serve as models for communities across the country.

[For highlights from the Rebuilding Strategy, click here.](#)

[To read the entire Rebuilding Strategy, click here.](#)



GET HELP OR GET INVOLVED

Help Others



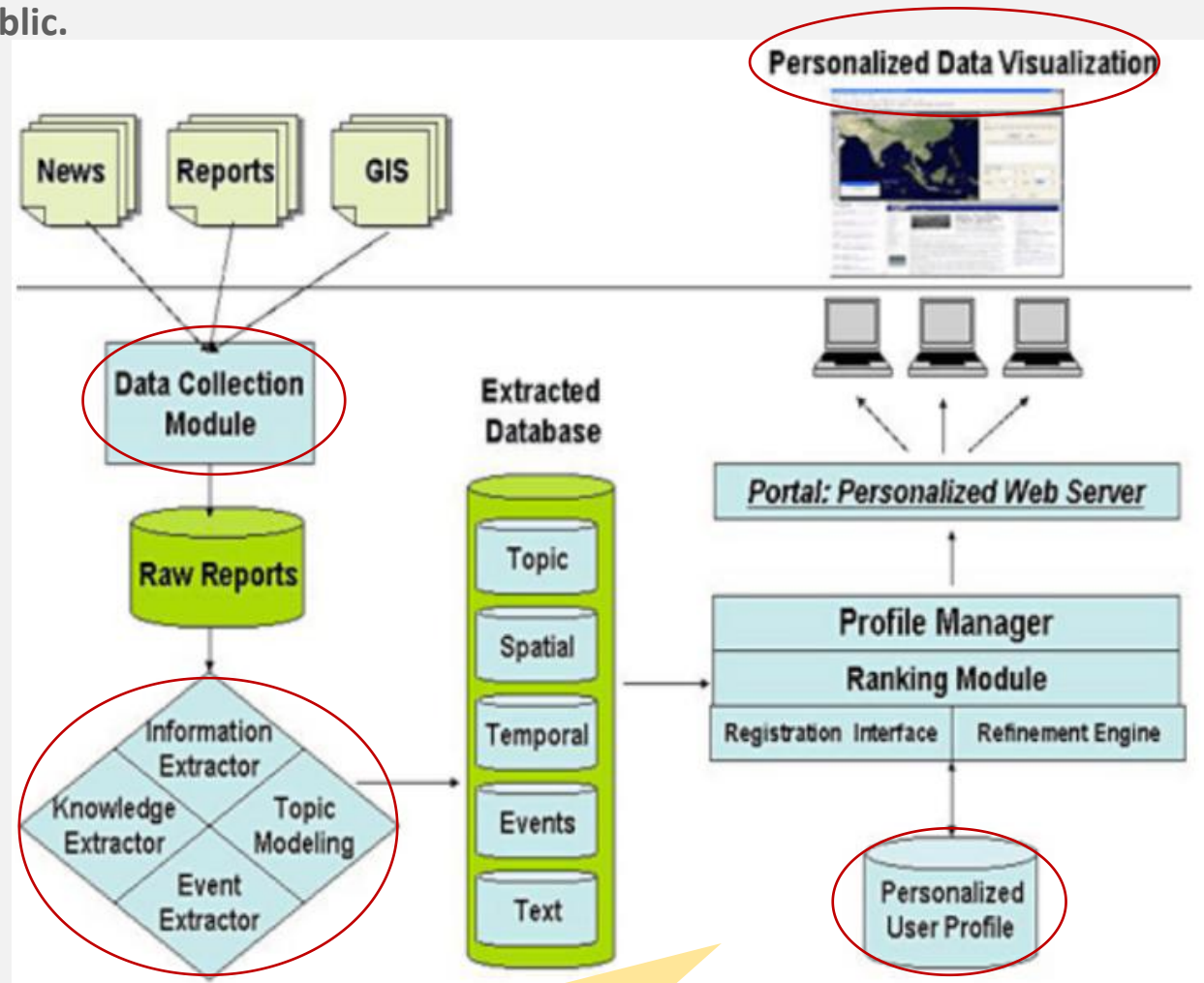
Get Help



➡ Current Situation and Problems

- **RESCUE Project:** funded by the National Science Foundation(NSF). Its goal is to dramatically improve the ability of emergency responders to gather, process, and disseminate information with each other and the general public.

—**For example,** PSAP (Personalized Situation Awareness Portal), an information portal, a research sponsored by RESCUE Project

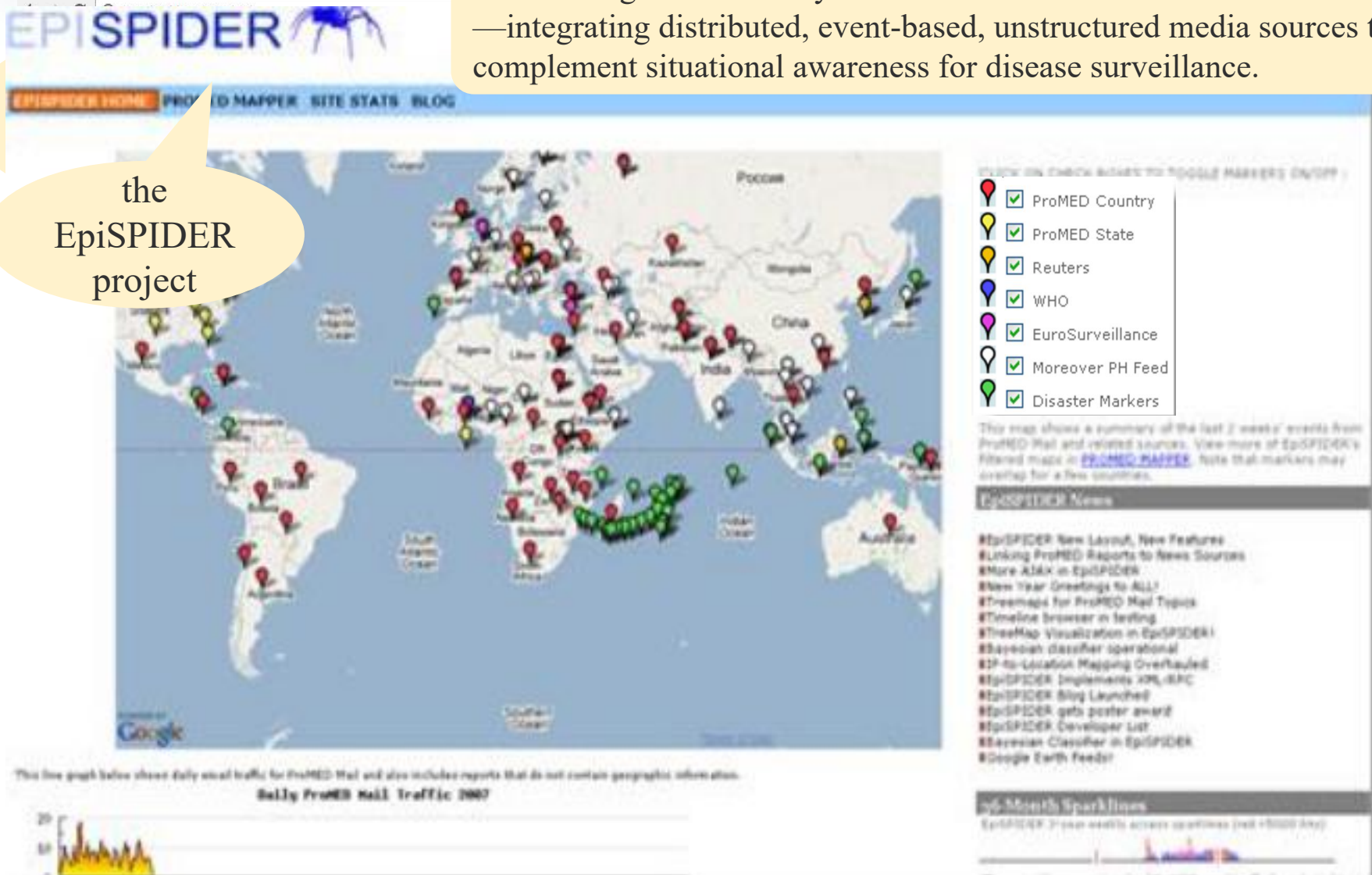


System Architecture of PSAP, illustrates the key technical components and how to connect them as an integrated system.

👉 Current Situation and Problems

- was designed in January 2006 .
- integrating distributed, event-based, unstructured media sources to complement situational awareness for disease surveillance.

the
EpiSPIDER
project



☞ Current Situation and Problems

Problems

The massive, heterogeneous and dynamic characteristics of Internet Information brings difficulties in information retrieval technology.

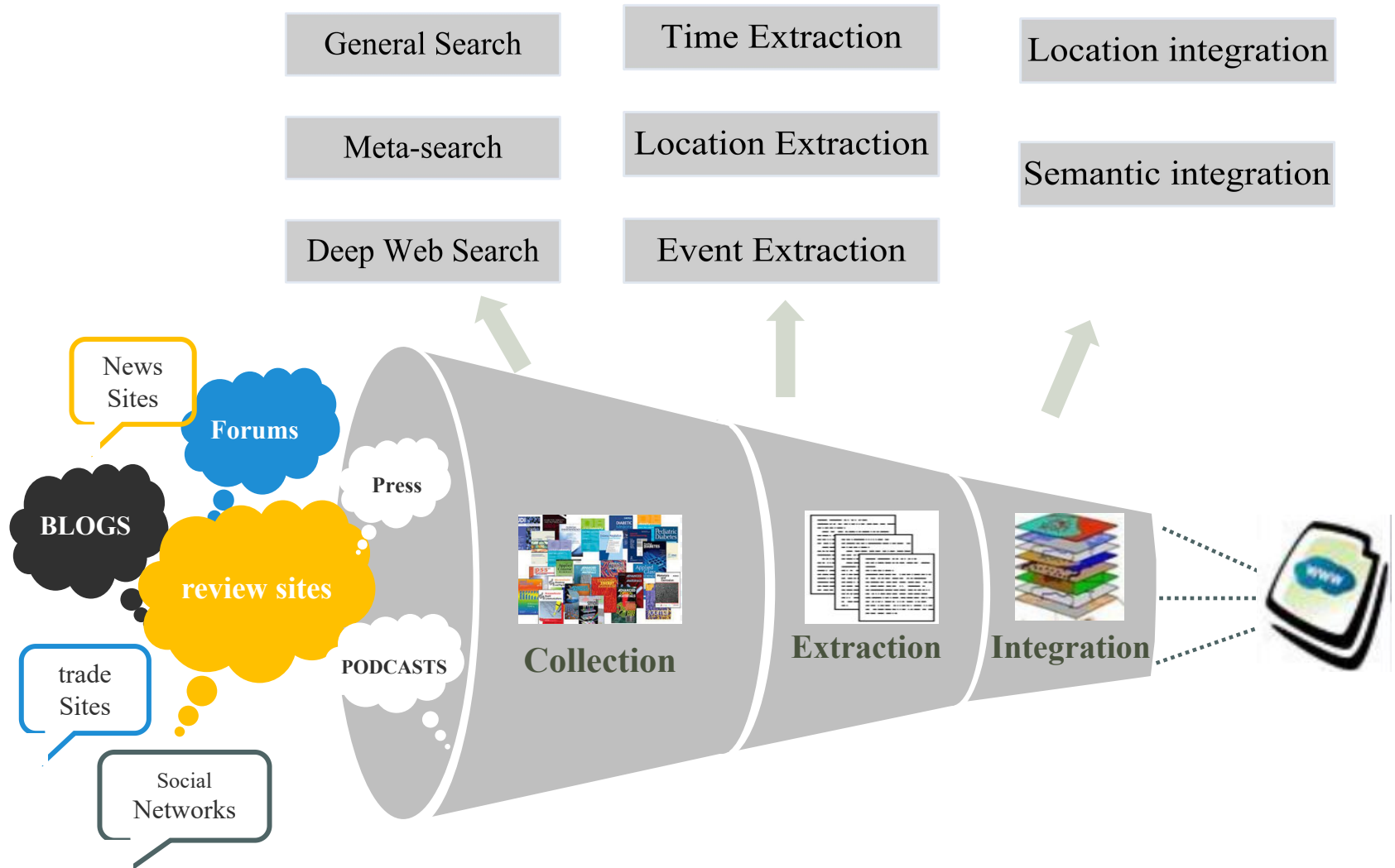
—**Spatial information search:** how to choose the driven crawler to get the ideal set of target sites.

—**Information Extraction:** how to distinguish and extract the polymorphic geographic information in web pages.

—**Information Integration:** how to deal with the Web data and apply them to practical application more appropriately.

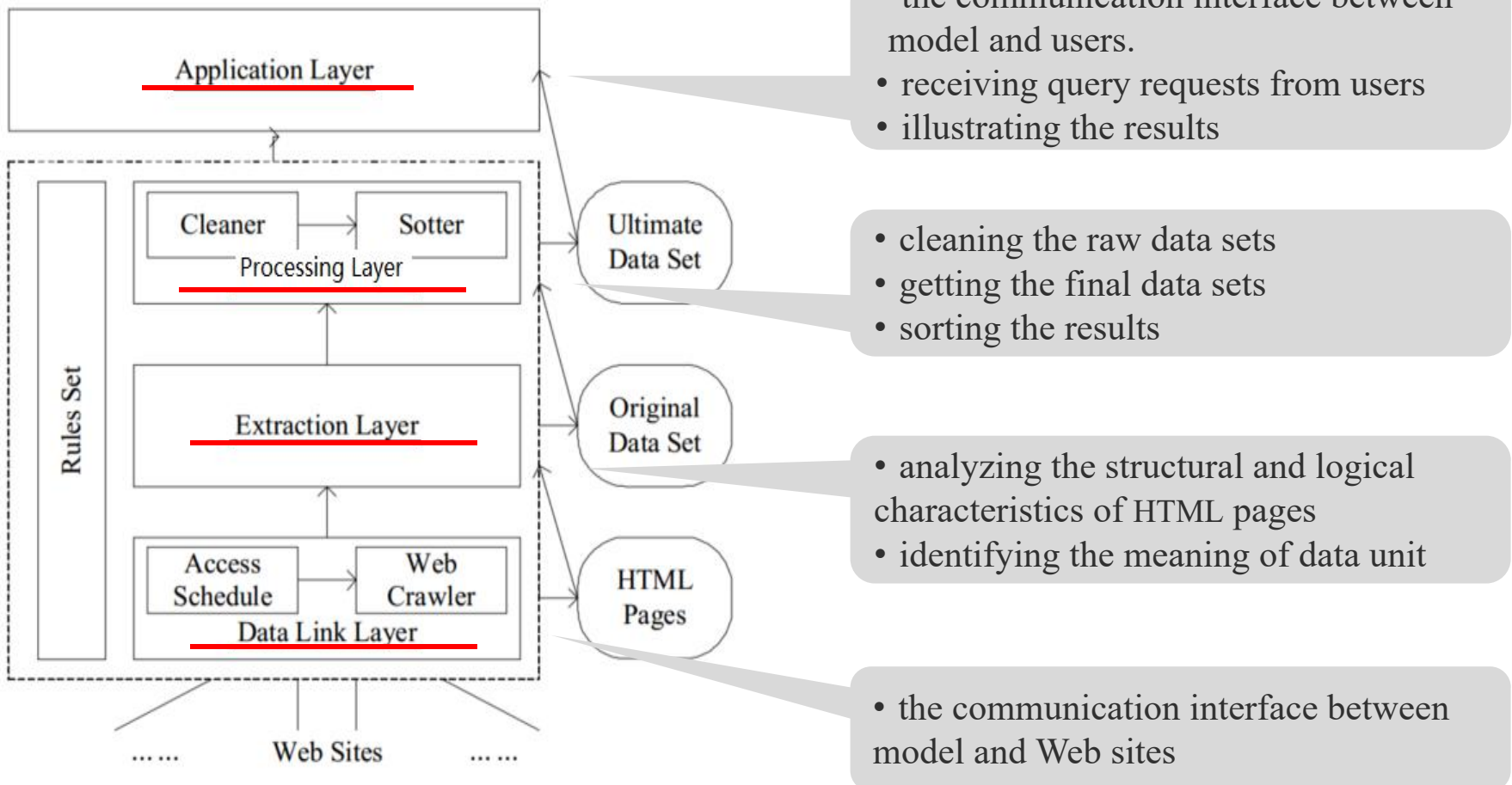
Research methods

Framework



👉 Research methods

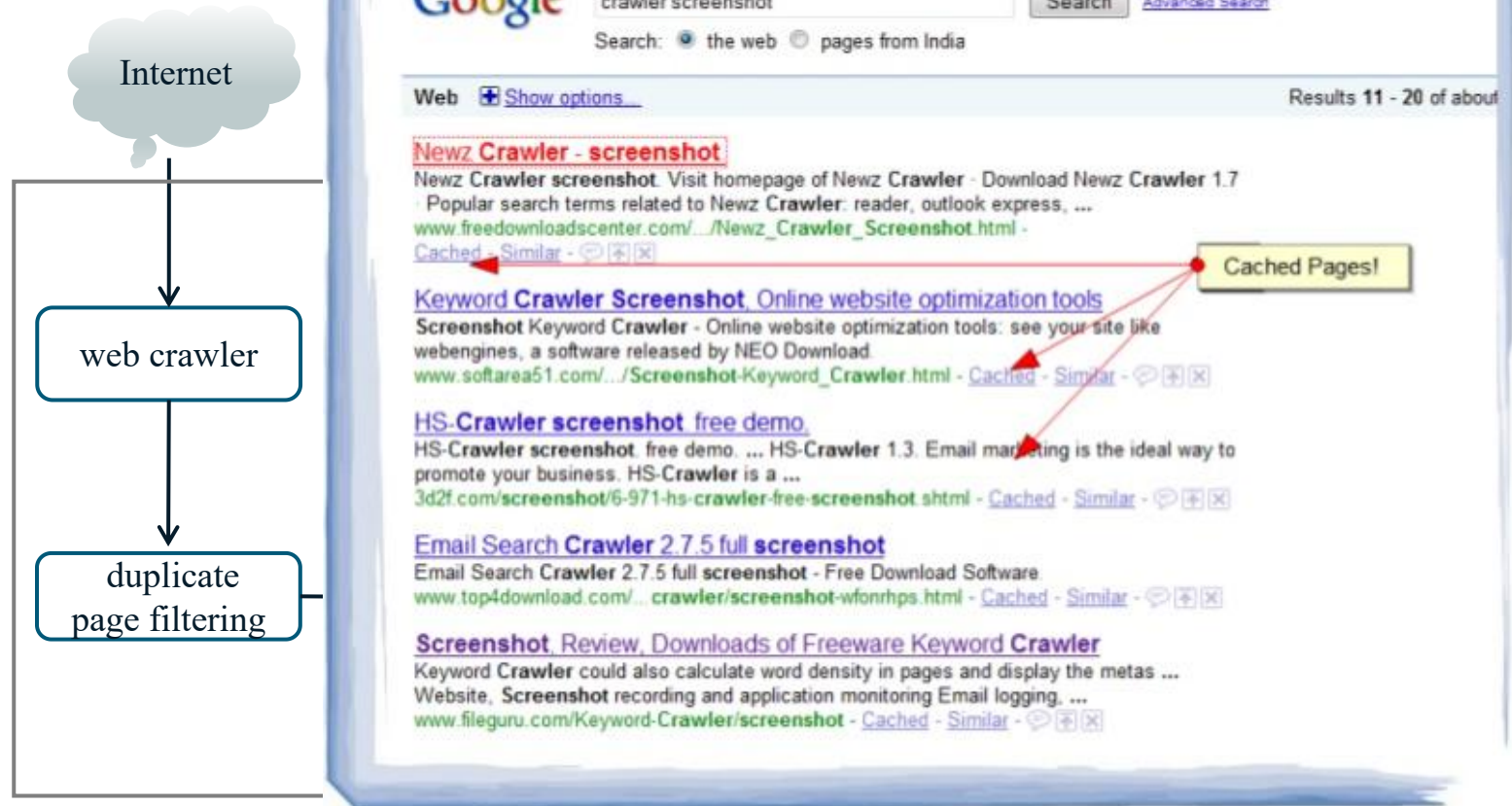
- The structure contains four main levels from bottom to top:



👉 Research methods——Collection

➤ General search

- an Internet portal through which a person can search numerous compiled resources for topic-related information.
- include **Google Yahoo! Alta Vista and Bing**
- to perform a search

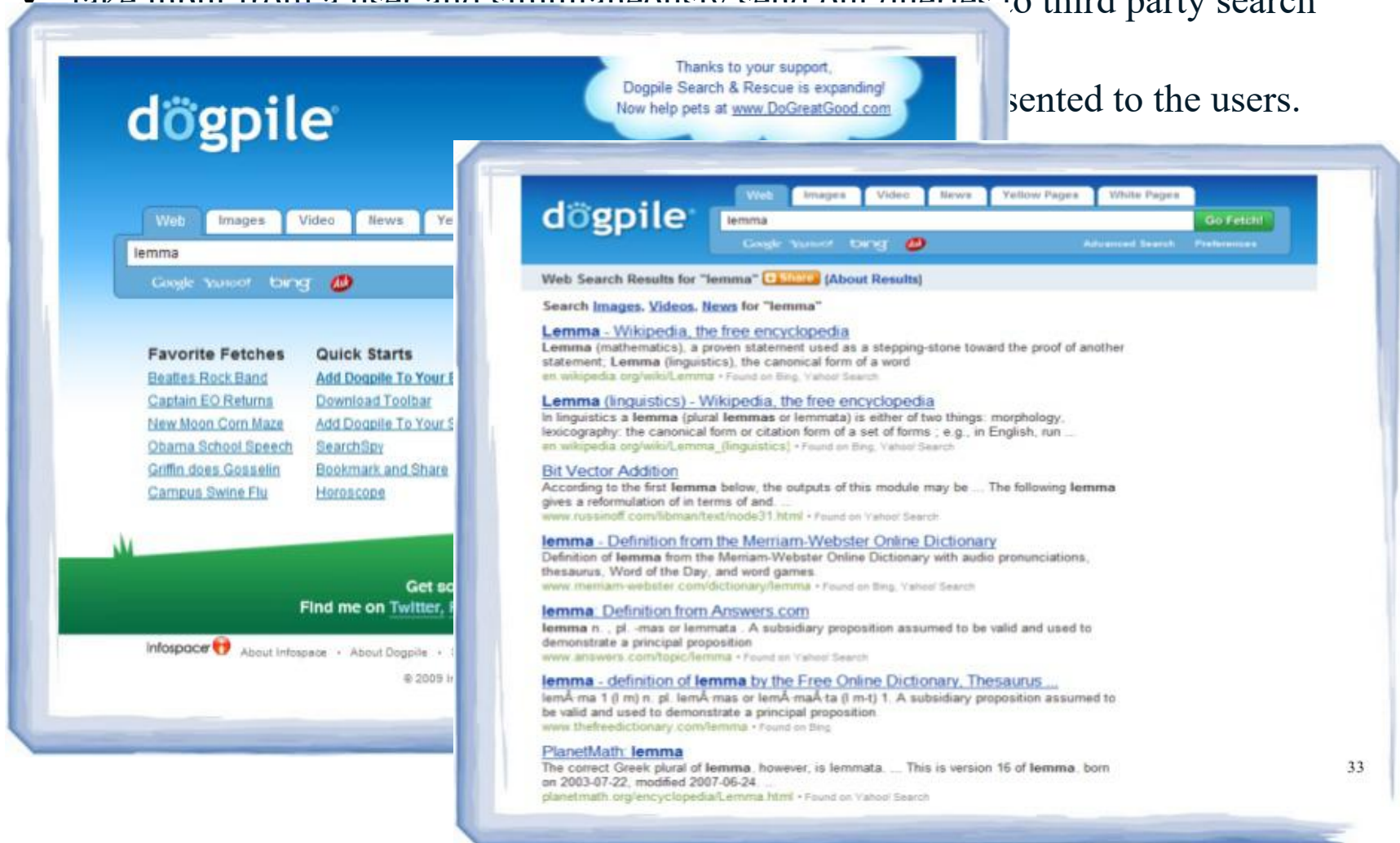


👉 Research methods—Collection

➤ Meta-search Engines

- a search tool that uses another search engine's data to produce their own results.
- take input from a user and simultaneously send out queries to third party search

presented to the users.



➡ Research methods——Collection

➤ Deep web search

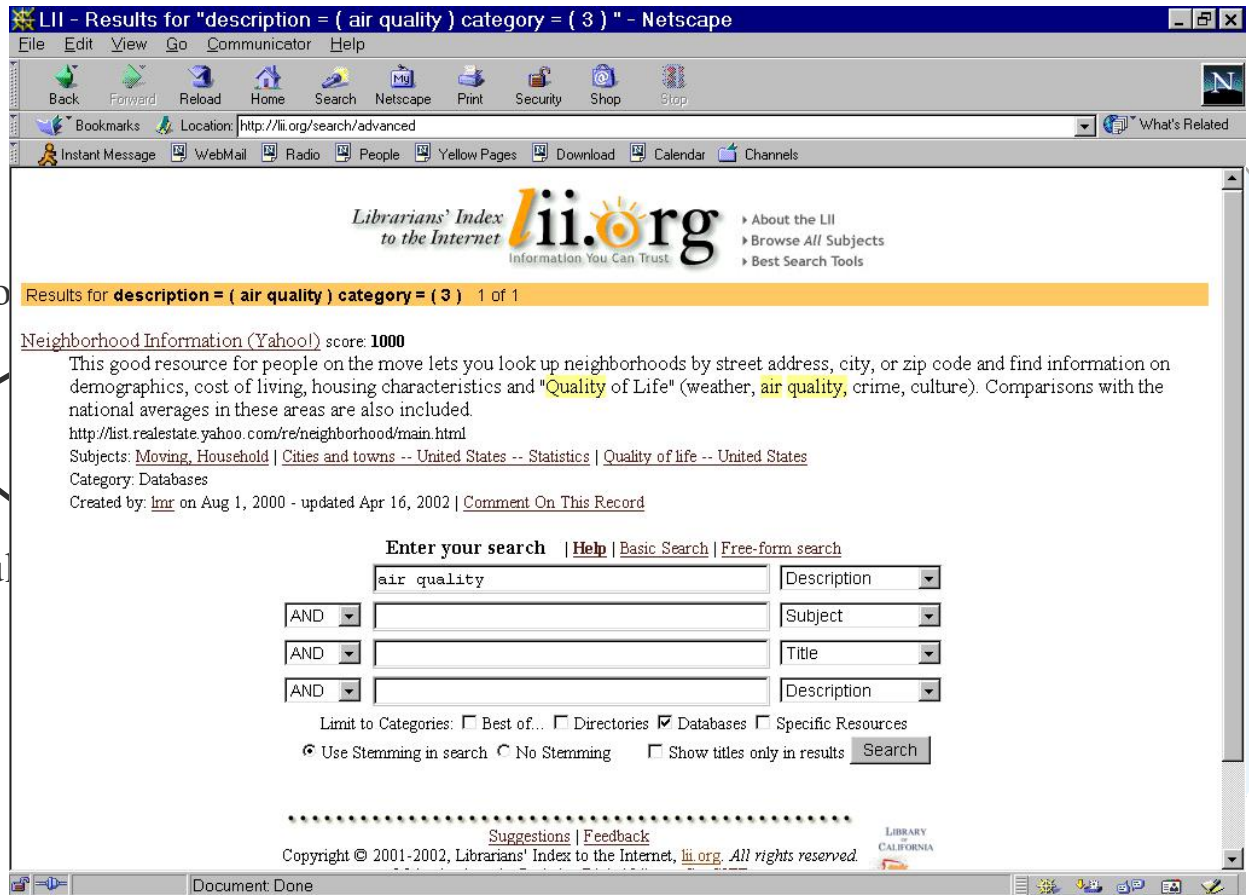
| Not Found by Standard Search Engines | Examples |
|--|--|
| Database content (dynamically generated for a particular inquiry) | Databases such as ERIC, PubMed, Library catalogs |
| Formats | pdf, audio, video, any new format |
| Real-time information | Current weather, stock quotations, airline flight checkers |
| Subscription or fee based services | EBSOChost, INFOTRAC, LexisNexis Academic |
| Sites requiring password access or registration for use | Blackboard, membership sites |
| Sites that require that forms be filled out | Travel direction sites, job hunting sites |
| Sites with a no-index protocol | Private sites |

- Deep Web—refers to all web pages that search engines cannot find,
- —such as user databases, registration-required web forums, webmail pages, and pages behind paywalls.
- —content behind web search forms on publicly available pages

➡ Research methods——Collection

➤ Deep web search

Deep web technologies: aims to achieve web data integration, by automating the process of analyzing database structures and cross referencing the results.



fill in a fo

Results for **description = (air quality) category = (3)** 1 of 1

Neighborhood Information (Yahoo!) score: **1000**

This good resource for people on the move lets you look up neighborhoods by street address, city, or zip code and find information on demographics, cost of living, housing characteristics and "Quality of Life" (weather, **air quality**, crime, culture). Comparisons with the national averages in these areas are also included.
<http://list.realestate.yahoo.com/re/neighborhood/main.html>
Subjects: [Moving, Household](#) | [Cities and towns -- United States -- Statistics](#) | [Quality of life -- United States](#)
Category: Databases
Created by: [lmr](#) on Aug 1, 2000 - updated Apr 16, 2002 | [Comment On This Record](#)

Enter your search | [Help](#) | [Basic Search](#) | [Free-form search](#)

| | |
|-------------|-------------|
| air quality | Description |
| AND | Subject |
| AND | Title |
| AND | Description |

Limit to Categories: ☐ Best of... ☐ Directories ☒ Databases ☐ Specific Resources
☒ Use Stemming in search ☐ No Stemming ☐ Show titles only in results

.....
[Suggestions](#) | [Feedback](#)
Copyright © 2001-2002, Librarians' Index to the Internet, [lii.org](#). All rights reserved.

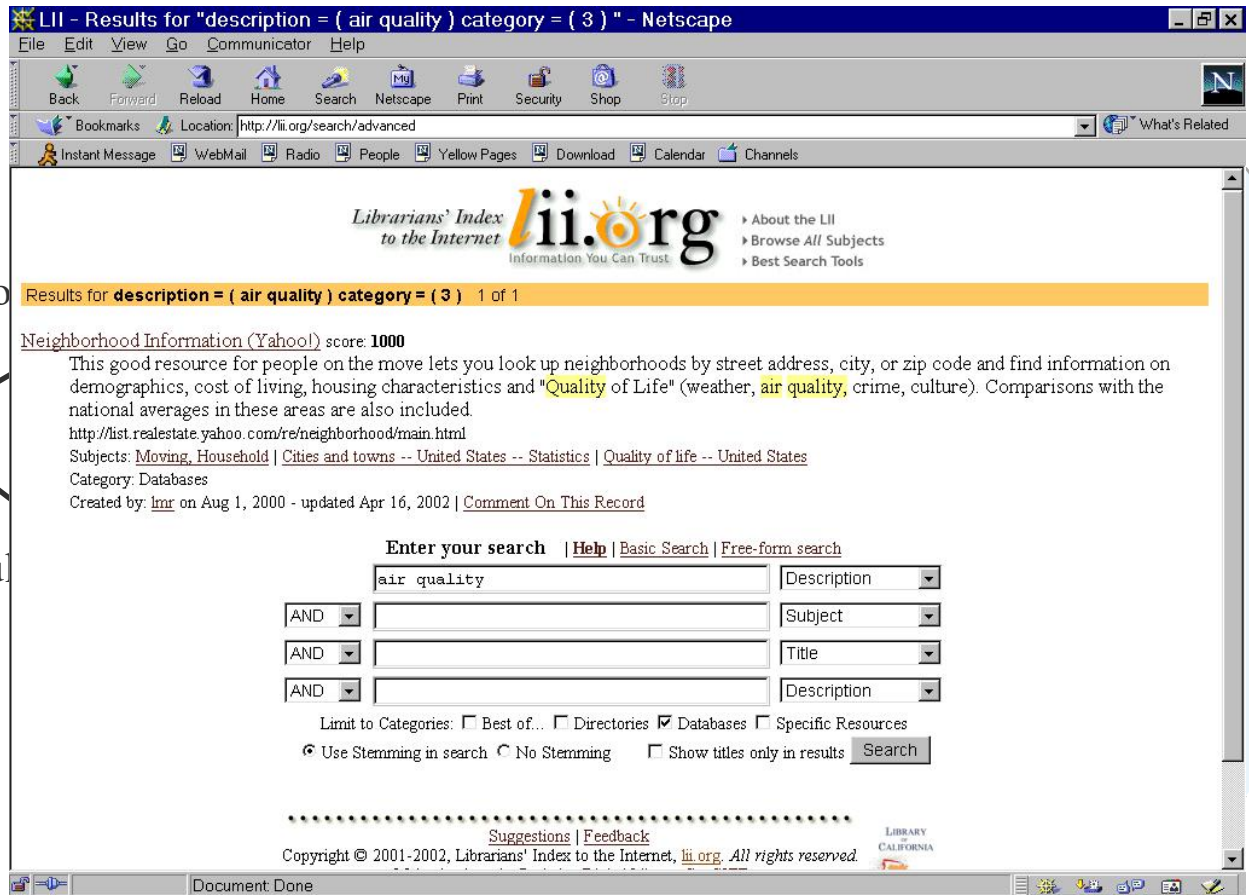
Document Done

Browse resu

➡ Research methods——Collection

➤ Deep web search

Deep web technologies: aims to achieve web data integration, by automating the process of analyzing database structures and cross referencing the results.



fill in a fo

Browse resu

Results for **description = (air quality) category = (3)** 1 of 1

Neighborhood Information (Yahoo!) score: **1000**

This good resource for people on the move lets you look up neighborhoods by street address, city, or zip code and find information on demographics, cost of living, housing characteristics and "Quality of Life" (weather, **air quality**, crime, culture). Comparisons with the national averages in these areas are also included.

<http://list.realestate.yahoo.com/re/neighborhood/main.html>

Subjects: [Moving, Household](#) | [Cities and towns -- United States -- Statistics](#) | [Quality of life -- United States](#)

Category: Databases

Created by: [lmr](#) on Aug 1, 2000 - updated Apr 16, 2002 | [Comment On This Record](#)

Enter your search | [Help](#) | [Basic Search](#) | [Free-form search](#)

air quality Description

AND Subject

AND Title

AND Description

Limit to Categories: ☐ Best of... ☐ Directories ☒ Databases ☐ Specific Resources

☒ Use Stemming in search ☐ No Stemming ☐ Show titles only in results

.....

[Suggestions](#) | [Feedback](#)

Copyright © 2001-2002, Librarians' Index to the Internet, [lii.org](#). All rights reserved.

LIBRARY OF CALIFORNIA

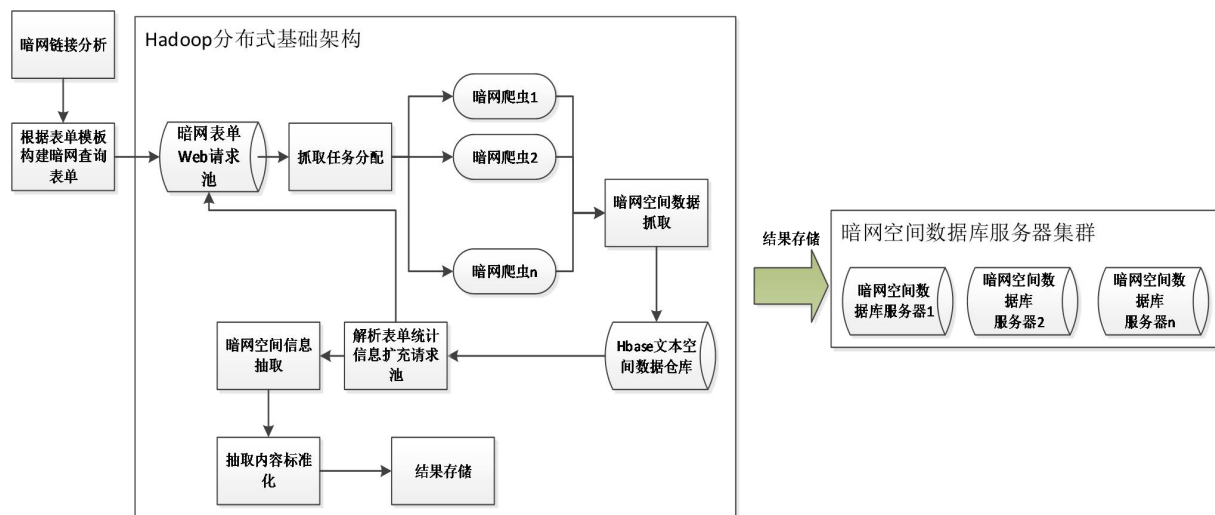
👉 Research methods——Collection

关键技术

➤ Deep web search

1、分布式DeepWeb空间信息采集

针对DeepWeb的抓取特点，利用Map/Reduce分布式系统架构、Web请求池构建、抓取任务弹性分配技术、异步I/O模型等关键技术，提出并实现暗网空间数据采集。

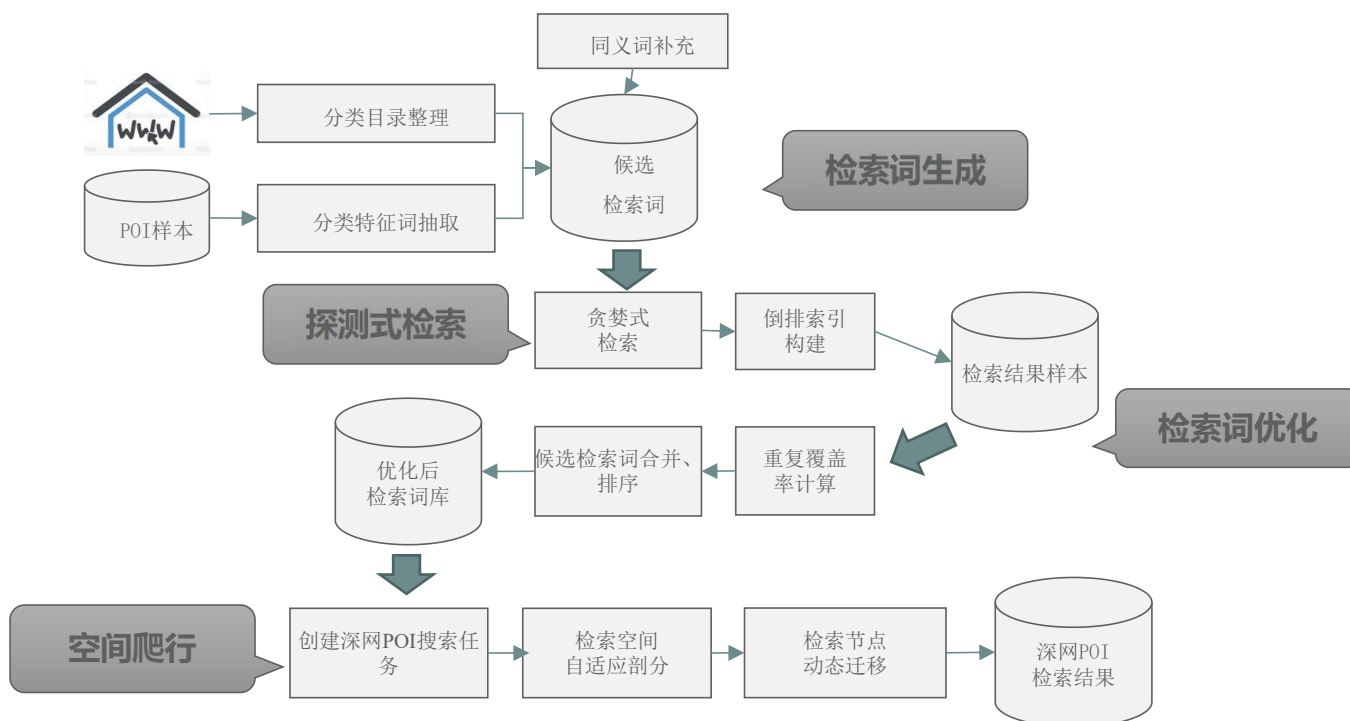


Research methods——Collection

关键技术

➤ Deep web search

1、分布式DeepWeb空间信息采集

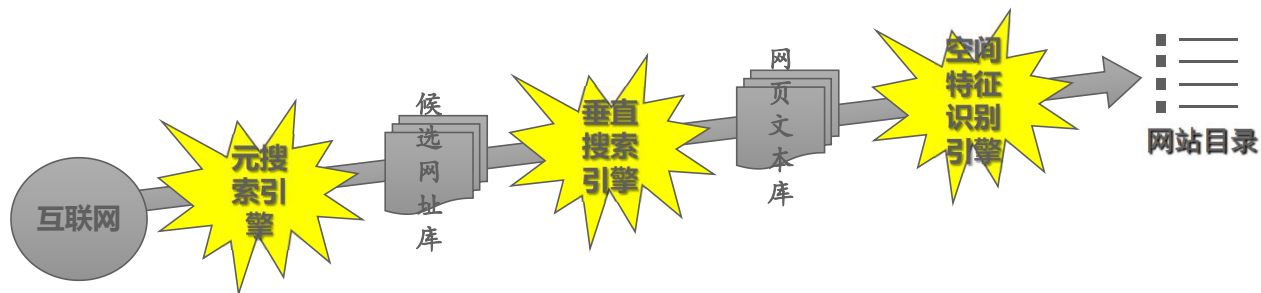


👉 Research methods——Collection

➤ Deep web search

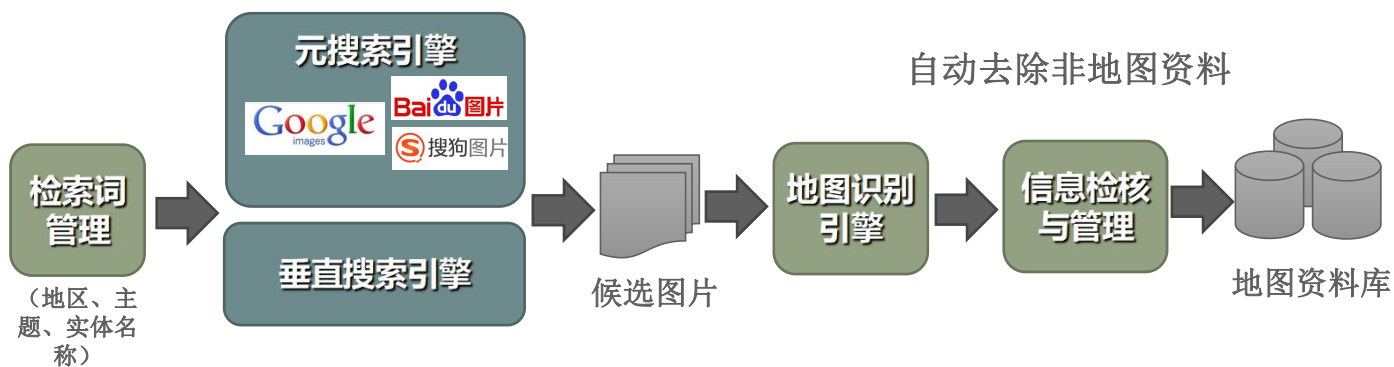
2、地理信息网站特征识别与自动发现

利用垂直搜索、元搜索等关键技术，分析和提取地理信息网站语义特征，构建地图网站发现与判别模型，实现地图网站的自动搜集与提取。



自动发现地理信息网站

3、基于像素统计与纹理特征的地图图像搜索



关键技术

3、基于像素统计与纹理特征的地图图像搜索

选择地图图片样本和非地图图片样本



提取样本特征，并标记类别，组成特征库



训练样本特征库，创建SVM(RBF)分类器



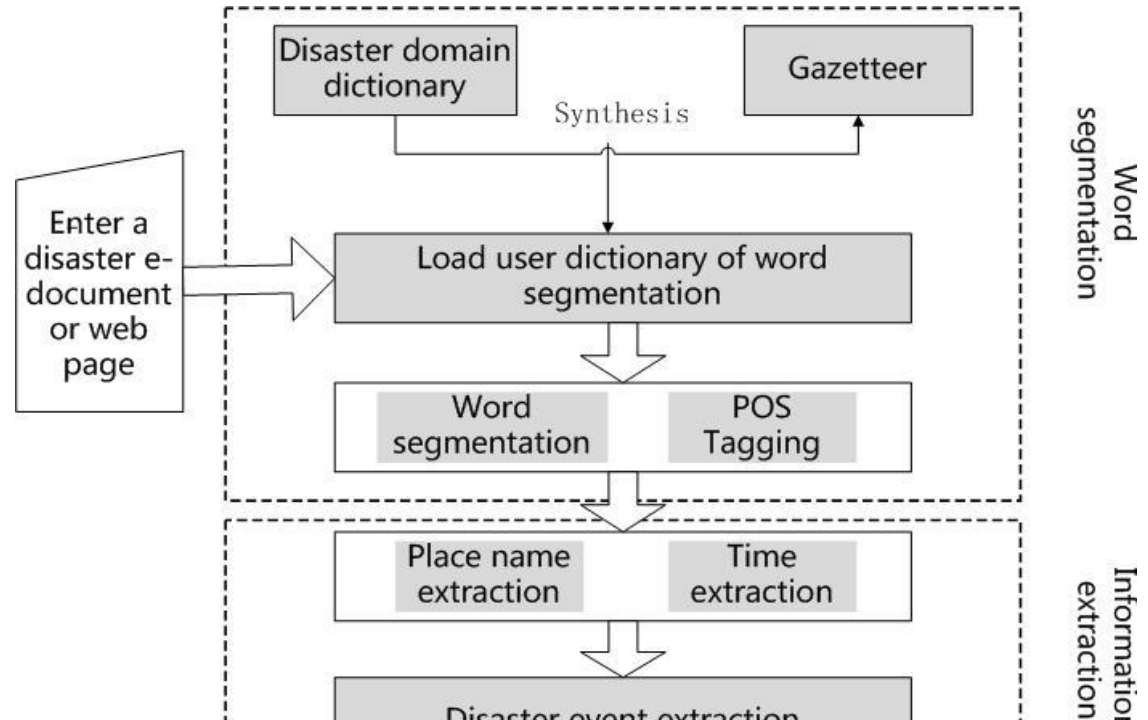
输入图片 提取特征 类别判定

矢量地图的自动识别准确率87%

👉 Research methods——Information Extraction

➤ Information Extraction:

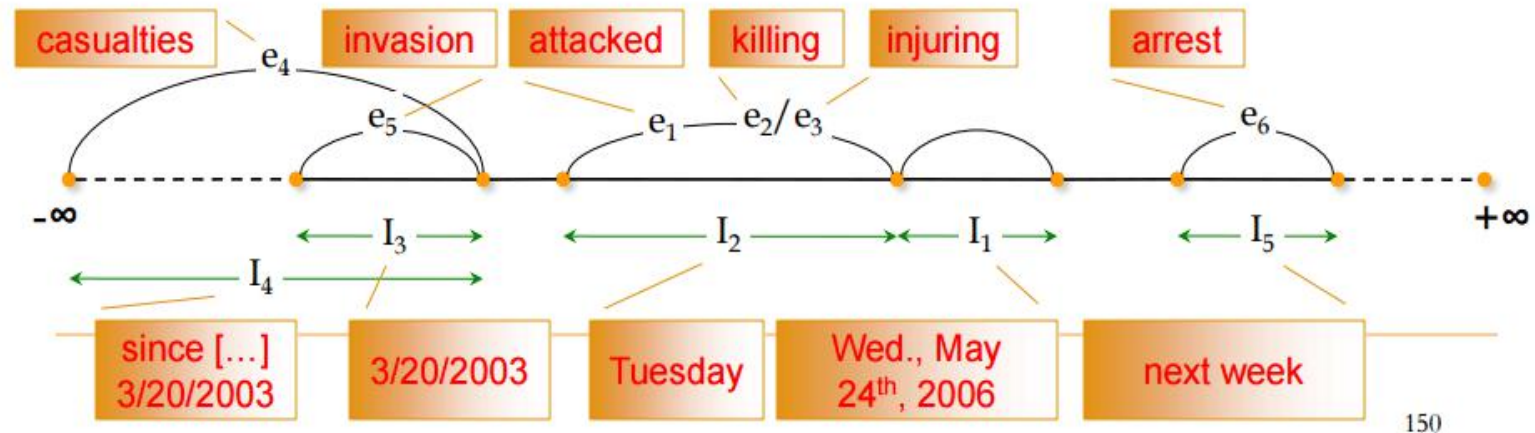
- Information Extraction (IE) is the process of finding structured text from unstructured or semi-structured text by annotating semantic information.
- **Knowledge base:** Lexical database, Syntactic pattern, Semantic relation
- **Information extraction model:** Rule models, Machine learning models
- **Semantic parsing model:** Spatial matching, space-time matching, space-time-attribute matching, space-time-attribute-event matching,



👉 Research methods——Information Extraction

➤ Time Extraction :

- based on the combination of trigger word and rule model
- information extraction and filtering process (three levels):
 1. time semantic role tagging
 2. phrase recognition
 3. syntactic pattern matching



150

Interval Based Event Timeline Construction

☞ Research methods——Information Extraction

➤ Location Extraction

- Place name extraction:
word segmentation and POS tagging
place name recognition
- Place name reasoning:
principle: the nearest distance

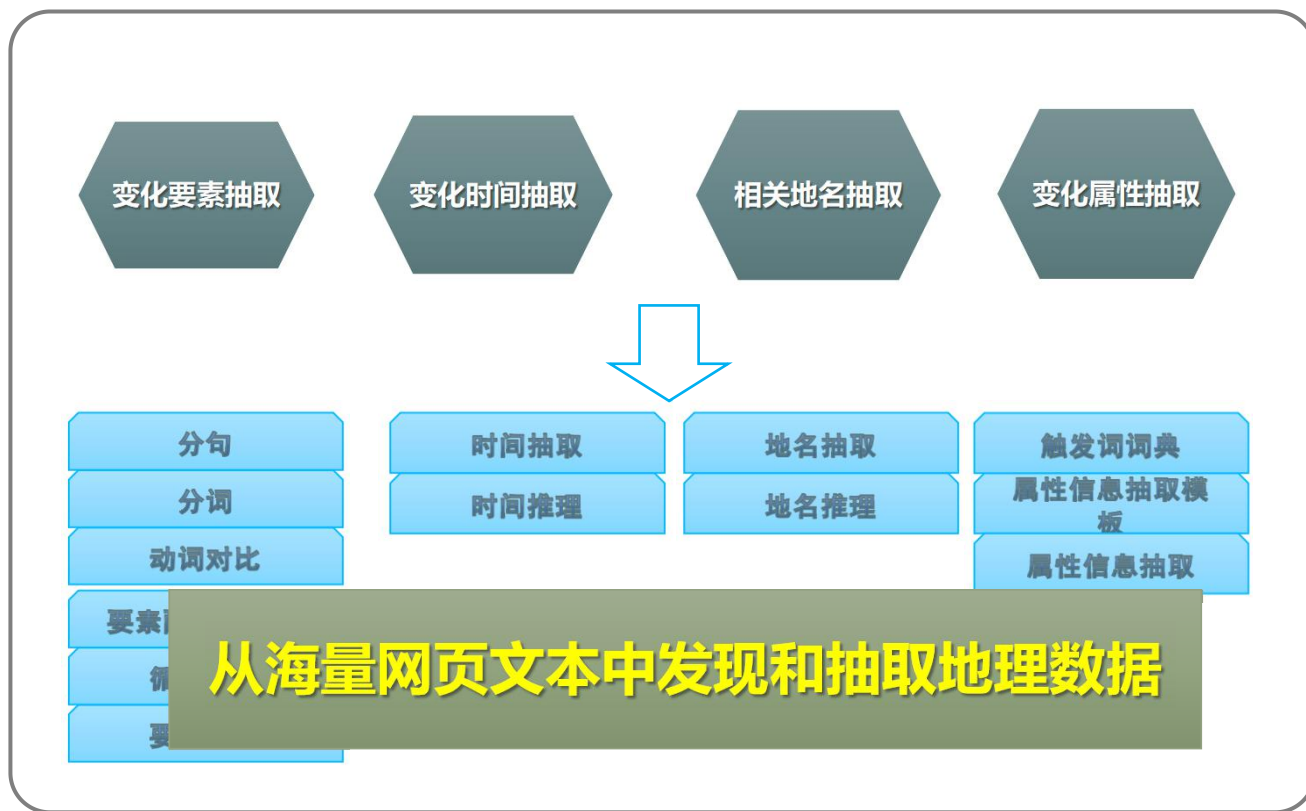
➤ Event Extraction

- Rule Learning based Approach:
dictionary based method, rule based method, wrapper induction.
- Classification Model based Approach:
Support Vector Machines (SVMs)
- Sequence Labeling based Approach:
Hidden Markov Models (HMMs) , Maximum Entropy Markov Models (MEMMs) ,
Conditional Random Fields (CRFs)

👉 Research methods——Information Extraction

➤ Geographical Entity Extraction:

关键技术

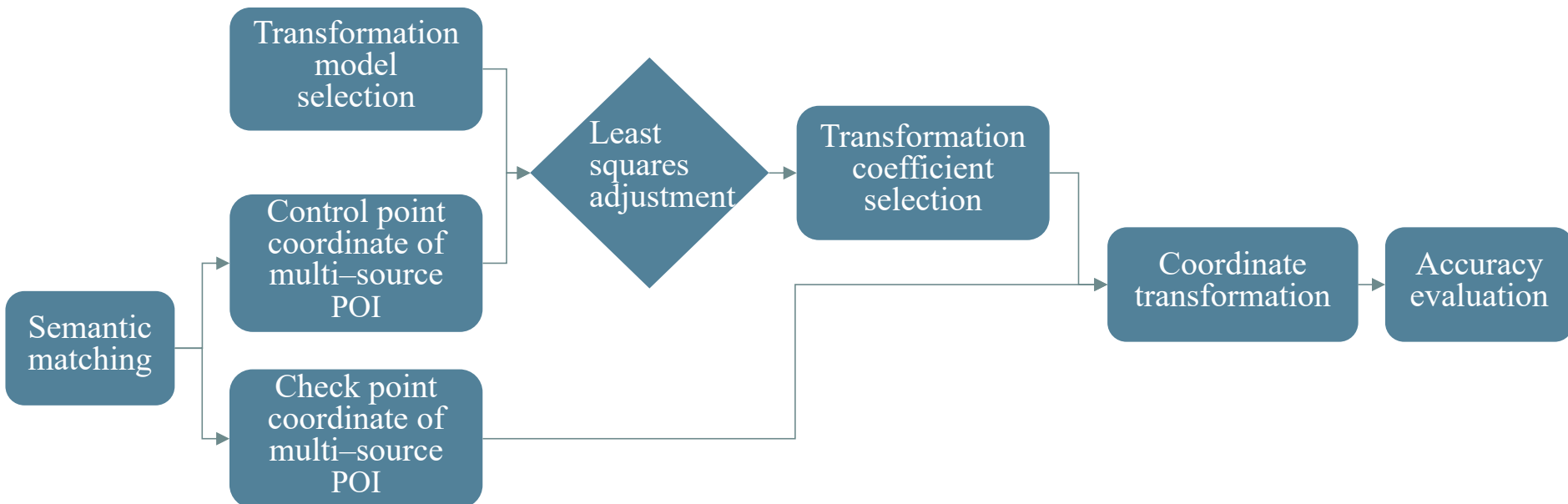


👉 Research methods——Information integration

➤ Location Integration——Spatial position correction of multi-source POI

There exists systematic deviation in the spatial position of different POI data.

- Obtaining the coordinate information of the control point
- Calculating the parameters of the coordinate transformation model
- Evaluating the spatial position correction accuracy.



Semantic-aided spatial location correction scheme for multi-source POI

👉 Research methods——Information integration

➤ Semantic Integration

- bring together information from diverse sources
- link them by using semantic information
- **three dimensions of semantic-integration research:**

Mapping discovery

- Using a Shared Ontology
- Using Heuristics and Machine-learning

Representation of mappings

- representing mappings as instances in an ontology of mappings
- using views to describe mappings

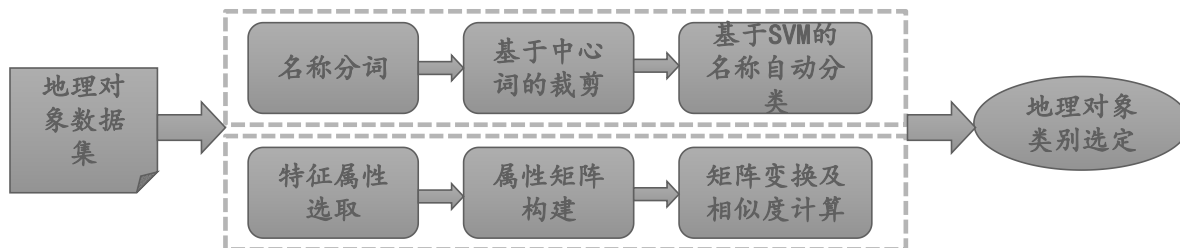
Reasoning with mappings

- reasoning over the source ontologies and the mappings

关键技术

4、地理对象语义分类

开展了兼顾文本与空间特征属性的地理对象语义分类技术研究，利用中文文本处理、语义相似度计算、空间分析等关键技术，实现基于角色标注的地理对象语义分类。

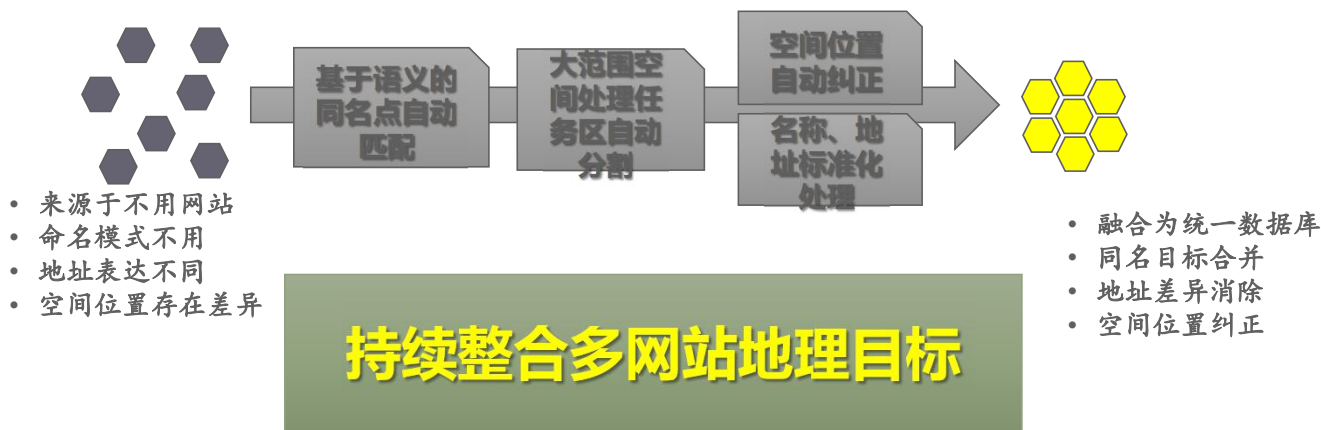


自动分类+跨网站类别转换

关键技术

5、多源地理信息数据融合处理

根据要素名称、描述信息、地理坐标等特征，建立统一的资源分类体系，实现多源信息的**名称消歧**和**地址信息标准化**；建立非线性坐标映射转换模型，实现**多源POI数据位置纠正**。



Typical experiments

Experimental research

- ✓ **Geographical Information Acquisition and Processing from Deep Web**
- ✓ **Background Map Search System**
- ✓ **Internet-based Information Extraction**
- ✓ **Integration of the Fundamental Geographical Information on the Internet**

Typical experiments

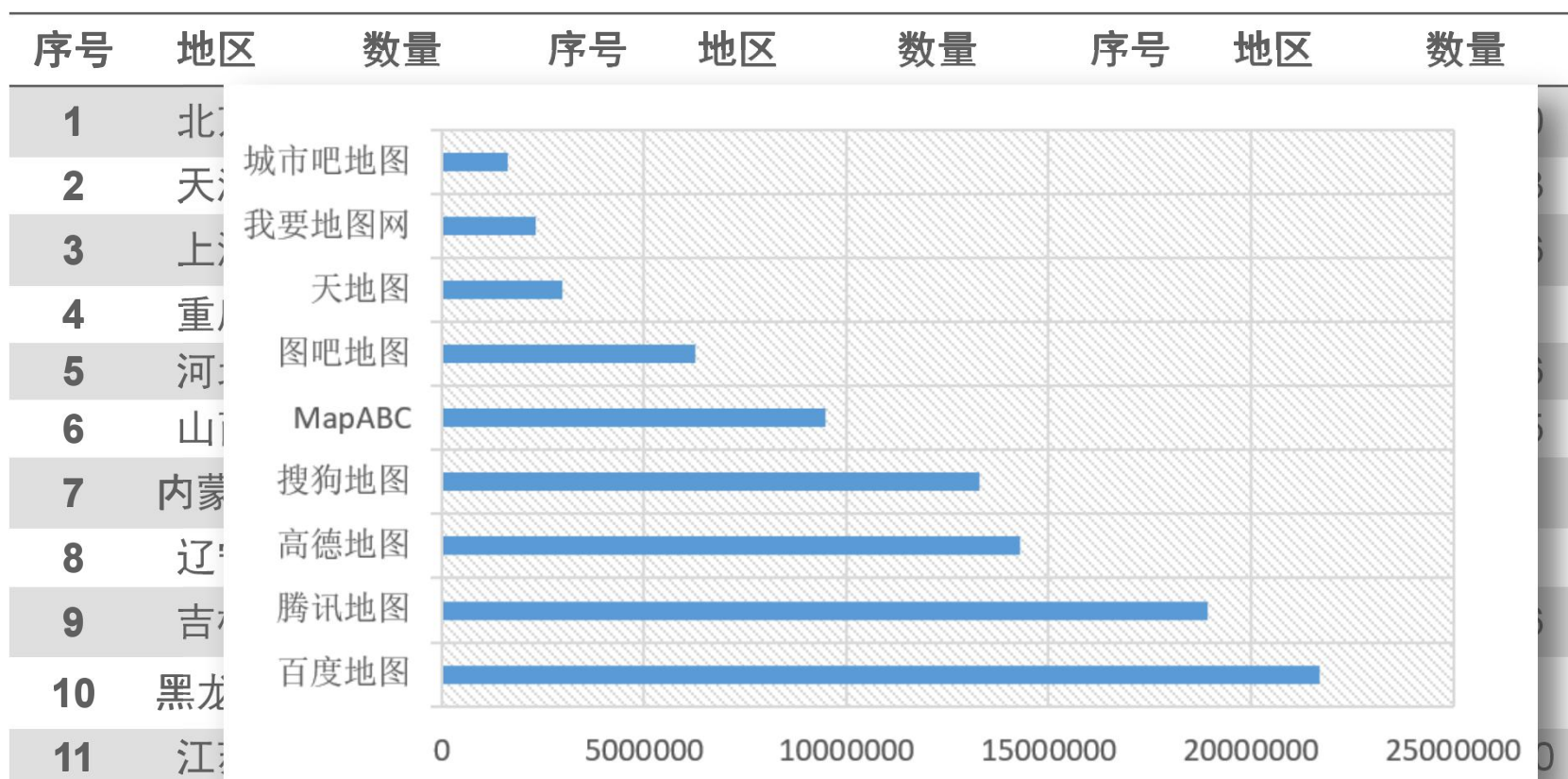
- Geographic Information Acquisition and processing from deep web pages
 - POI search engine:



Typical experiments

➤ Geographic Information Acquisition and processing from deep web pages

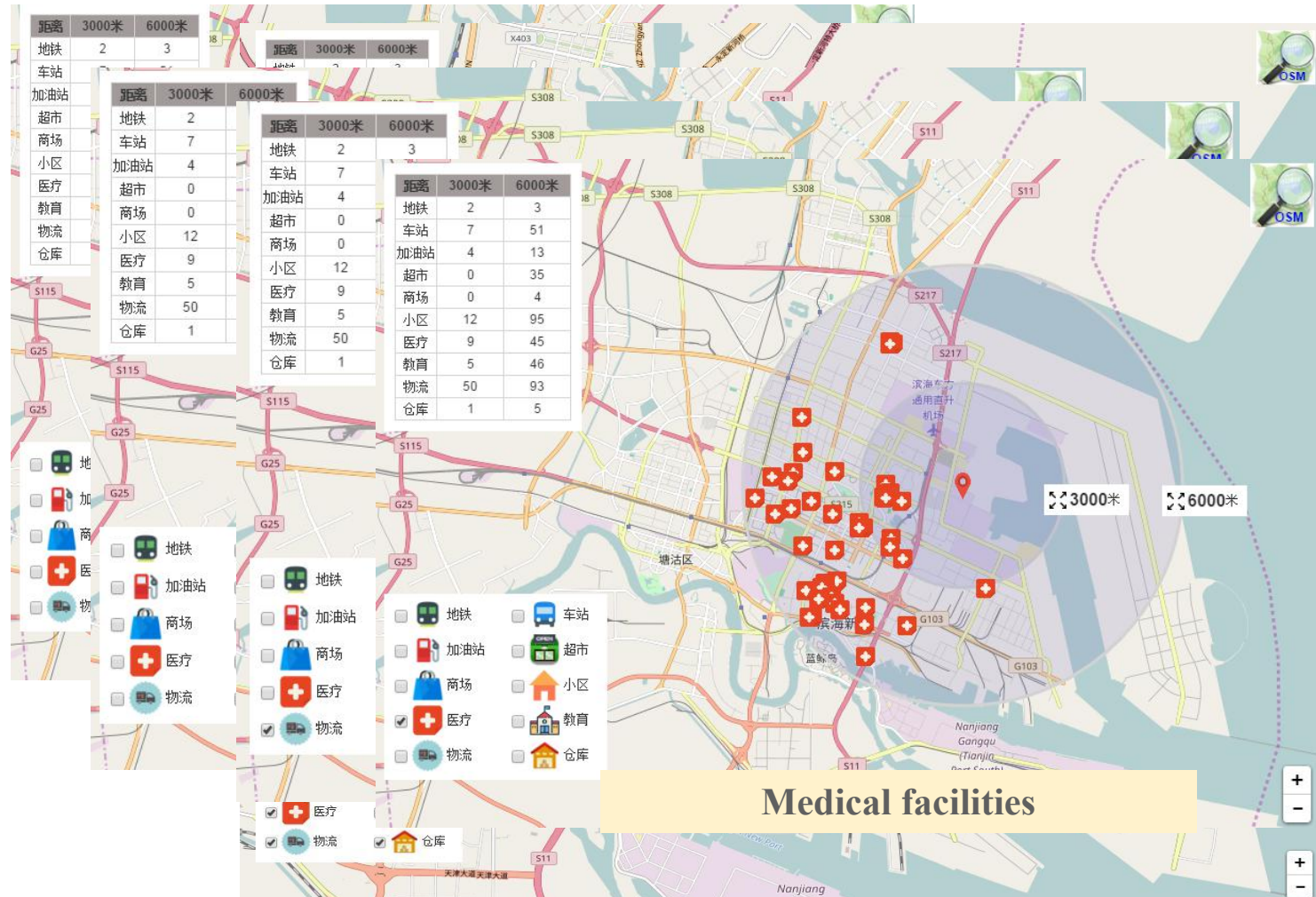
Statistics on the number of POI in different provinces:



- Crawling search test was conducted on the 9 kinds of target website.
- Obtained a total of about 90.87 million POI (except Taiwan) within three months.
- POI number of Baidu, Gaode, Tencent and Sogou is respectively more than 10 million.

Typical experiments

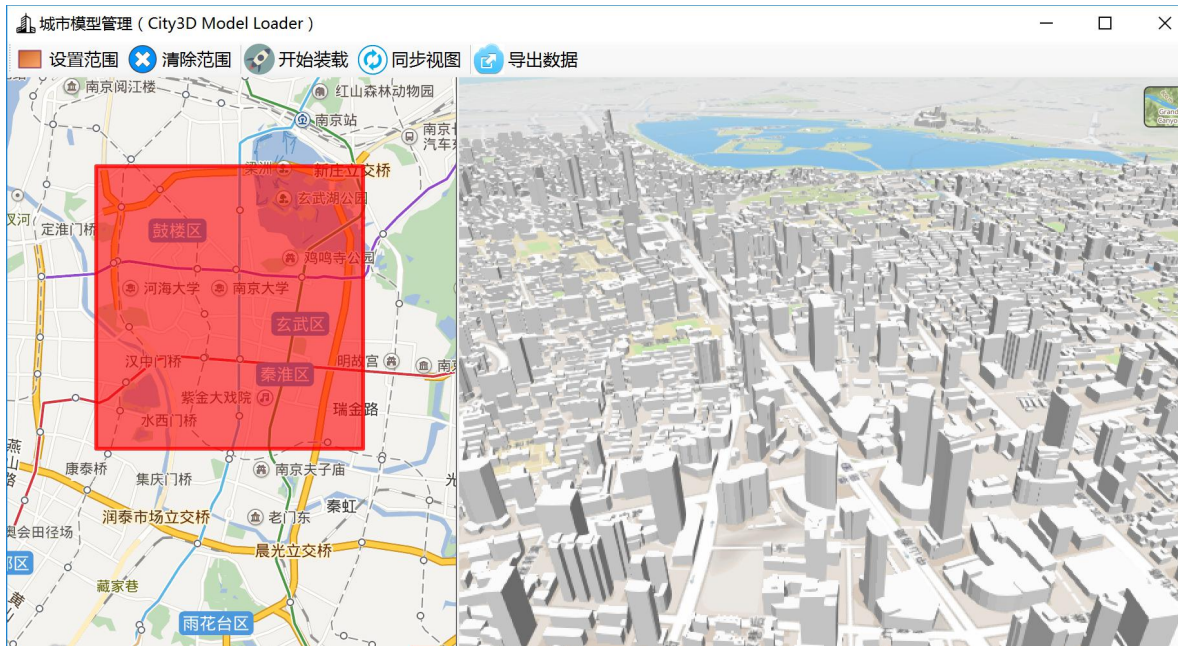
The distribution of POI in 3km and 6km around the site of explosion in Binhai New Area, Tianjin



Typical experiments

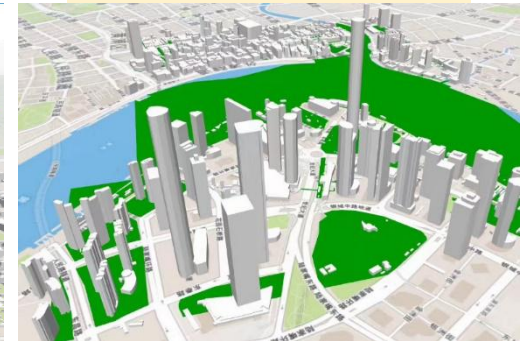
关键技术

- Geographic Information Acquisition and processing from deep web pages
 - Urban construction information acquisition:



Nanjing

Shanghai



Beijing

Typical experiments

• Disaster events search:

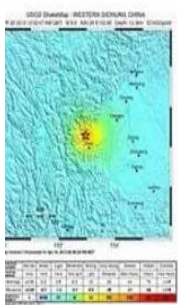
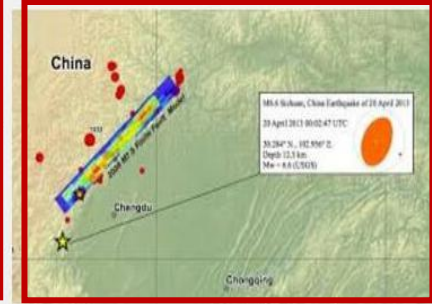
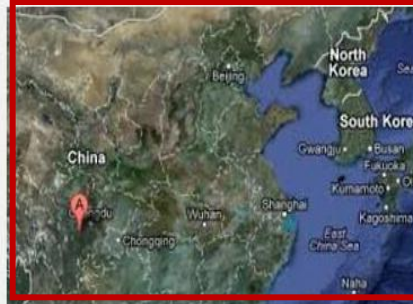
| event name | content | time | URL | Area | Location | |
|-----------------|-------------------------------------|--------------|---------------------|---------------------------------------|----------|-----------------------|
| event name | content | time | URL | Area | Location | |
| 海淀区发生一起1人死亡事故 | 12月22日15时30分,在海淀区温泉镇双坡路路北北京科弘投资管 | 12月22日15时30分 | 12/22/2015 15:30:00 | http://www.bjsafety.gov.cn/accidentin | 北京市海淀区 | 116.3218639.95606977 |
| 海淀区发生一起1人死亡事故 | 12月18日10时许,在海淀区清华大学化学系实验室发生一起爆炸 | 12月18日10时 | 12/18/2015 10:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市海淀区 | 116.3218639.95606977 |
| 顺义区发生一起1人死亡事故 | 12月18日8时50分,顺义区北京现代汽车有限公司一厂废料处理厂 | 12月18日8时50分 | 12/18/2015 08:50:00 | http://www.bjsafety.gov.cn/accidentin | 北京市顺义区 | 116.6679340.128491057 |
| 通州区六环一甲醇罐车遭拉煤车 | 12月17日9时15分,在通州区六环徐辛庄内环路,一拉煤货车道 | 12月17日9时15分 | 12/17/2015 09:15:00 | http://www.bjsafety.gov.cn/accidentin | 北京市通州区 | 116.6912839.896040779 |
| 西城区发生不明气体情况 | 12月3日中午,北京四中初中部周边有疑似汽油味道,西城区安监 | 12月3日 | 12/3/2015 00:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市西城区 | 116.3534539.8928799 |
| 海淀区发生一起1人死亡事故 | 11月16日8时许,在海淀区西三环花园桥东约200米路北,北京 | 11月16日8时 | 11/16/2015 08:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市海淀区 | 116.3218639.95606977 |
| 海淀西北旺发生一起坍塌事故 | 北京龙泉机动车驾驶员培训中心承租的西北旺甲1号空地南侧山 | 10月30日16时30分 | 10/30/2015 16:30:00 | http://www.bjsafety.gov.cn/accidentin | 北京市 | 116.3946539.894104004 |
| 门头沟区发生一起高处坠落事故 | 10月11日13时许,在北京中铁西城工程项目工地,工人聂某某(| 10月11日13时 | 10/11/2015 13:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市门头沟 | 116.0980239.932906923 |
| 海淀区发生一起高处坠落事故 | 10月11日14时50分,在中关村大街59号人大逸夫会堂室外设施 | 10月11日14时50分 | 10/11/2015 14:50:00 | http://www.bjsafety.gov.cn/accidentin | 北京市海淀区 | 116.3218639.95606977 |
| 海淀区发生一起工地砸伤1人死亡 | 10月11日11时40分,在海淀区清河毛纺路阳光南里32号楼南侧 | 10月11日11时40分 | 10/11/2015 11:40:00 | http://www.bjsafety.gov.cn/accidentin | 北京市海淀区 | 116.3218639.95606977 |
| 通州区发生一起一人死亡事故 | 10月6日16时许,北京岱摩斯变速器有限公司委托现代重工对本 | 10月6日16时7分 | 10/6/2015 16:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市通州区 | 116.6912839.896040779 |
| 平谷区发生一起一人死亡事故 | 10月3日16时许,在平谷区山东庄镇桥头营村西北京浩天荣华商 | 10月3日16时17分 | 10/3/2015 16:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市平谷区 | 117.1279540.137940577 |
| 大兴区发生一起2死2伤事故 | 10月1日10时20分,北京康家兴旺物业管理有限公司工人肖自叶、 | 10月1日10时20分 | 10/1/2015 10:20:00 | http://www.bjsafety.gov.cn/accidentin | 北京市大兴区 | 116.3520839.747321955 |
| 朝阳区发生一起1人死亡事故 | 10月1日13时30分许,在朝阳区建外金地中心A座,北京中建华宇机 | 10月1日13时30分 | 10/1/2015 13:30:00 | http://www.bjsafety.gov.cn/accidentin | 北京市朝阳区 | 116.4976539.919742721 |
| 朝阳区发生一起燃气泄漏事故 | 9月30日8时40分,在朝阳区东坝乡东坝南二街二期3号楼段市政 | 9月30日8时40分 | 9/30/2015 08:40:00 | http://www.bjsafety.gov.cn/accidentin | 北京市朝阳区 | 116.4976539.919742721 |
| 朝阳区发生一起1人死亡事故 | 9月12日13时许,北京市朝阳安贞地区的808工程工地,山东省肥 | 9月12日13时16分 | 9/12/2015 13:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市朝阳区 | 116.4976539.919742721 |
| 房山区发生一起一死一伤事故 | 9月5日18时10分许,北京中铁房山桥梁有限公司生产区轨枕车间 | 9月5日18时10分 | 9/5/2015 18:10:00 | http://www.bjsafety.gov.cn/accidentin | 北京市房山区 | 116.1735539.732537984 |
| 怀柔区发生一起1人死亡事故 | 8月18日1时30分,在怀柔区雁栖经济开发区北京威亚高性能纤维 | 8月18日1时30分 | 8/18/2015 01:30:00 | http://www.bjsafety.gov.cn/accidentin | 北京市怀柔区 | 116.669340.313043205 |
| 顺义区发生一起1人死亡事故 | 15日14时30分许,在顺义区南法信镇焦各庄村南30米,钩机司机 | 15日14时30分 | 8/15/2015 14:30:00 | http://www.bjsafety.gov.cn/accidentin | 北京市顺义区 | 116.6679340.128491057 |
| 朝阳区发生一起1人死亡事故 | 【事故信息】8月4日21时04分,接市应急办通报:8月3日16时18分 | 8月4日21时04分 | 8/4/2015 21:04:00 | http://www.bjsafety.gov.cn/accidentin | 北京市朝阳区 | 116.4976539.919742721 |
| 昌平区阳坊镇一饭馆煤气罐爆燃 | 7月29日7时38分,阳坊镇政府对面一驴肉火烧店(北京昌平幸福 | 7月29日7时38分 | 7/29/2015 07:38:00 | http://www.bjsafety.gov.cn/accidentin | 北京市昌平区 | 116.262840.216635475 |
| 海淀区发生一起触电死亡事故 | 7月28日14时50分左右,在海淀区中关村东路苏宁联想桥店,北 | 7月28日14时50分 | 7/28/2015 14:50:00 | http://www.bjsafety.gov.cn/accidentin | 北京市海淀区 | 116.3218639.95606977 |
| 大兴区发生一起火灾事故 | 7月19日14时10分许,在通州区梨园镇大福村北京市大福京苑加 | 7月19日14时10分 | 7/19/2015 14:10:00 | http://www.bjsafety.gov.cn/accidentin | 北京市通州区 | 116.6912839.896040779 |
| 通州区发生一起一人受伤事故 | 7月18日16时左右,在通州区台湖镇,东城区安置房地三层彩 | 7月18日16时 | 7/18/2015 16:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市东城区 | 116.4056339.887611445 |
| 大兴区发生一起火灾事故 | 13时20分,大兴区狼垡北京铁路局动车段院内,中铁七局下属的 | 13时20分 | 1/1/0001 00:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市大兴区 | 116.3520839.747321955 |
| 昌平区发生一起一人死亡事故 | 7月17日19时许,在昌平区阳坊阅兵训练基地北门外约1000米处 | 7月17日19时 | 7/17/2015 19:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市昌平区 | 116.262840.216635475 |
| 房山区发生一起氨水泄漏事故 | 7月15日20时50分,位于房山区燕化公司炼油三厂三蜡成型装置 | 7月15日20时50分 | 7/15/2015 20:50:00 | http://www.bjsafety.gov.cn/accidentin | 北京市房山区 | 116.1735539.732537984 |
| 丰台区发生一起火灾事故 | 7月12日16时13分,在丰台区长辛店南岗洼村村委会西侧有限公 | 7月12日16时13分 | 7/12/2015 16:13:00 | http://www.bjsafety.gov.cn/accidentin | 北京市丰台区 | 116.3184339.852302354 |
| 房山区发生一起交通事故 | 7月7日14时许,北京燕飞蓝天润滑油厂 | | | http://www.bjsafety.gov.cn/accidentin | 北京市房山区 | 116.1735539.732537984 |
| 西城区发生一起触电事故 | 7月6日15时左右,中国联通公司的施工单 | | | http://www.bjsafety.gov.cn/accidentin | 北京市西城区 | 116.3534539.8928799 |
| 朝阳区发生一起火灾事故 | 7月7日11时30分许,朝阳区建外街道永泰 | | | http://www.bjsafety.gov.cn/accidentin | 北京市朝阳区 | 116.4976539.919742721 |
| 密云县发生一起死亡事故 | 6月22日15时30分许,在密云县溪翁庄镇京 | | | http://www.bjsafety.gov.cn/accidentin | 北京市密云县 | 116.8615740.365381076 |
| 朝阳区发生一起高坠事故 | 11时30分许,朝阳区黑庄户乡双桥村280号在建库房,北京兴龙伴 | 11时30分 | 1/1/0001 00:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市朝阳区 | 116.4976539.919742721 |
| 丰台区发生一起煤气罐泄漏事故 | 17时15分,丰台区卢沟桥乡晓月园三里一海鲜饺子馆后厨液化气 | 17时15分 | 1/1/0001 00:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市丰台区 | 116.3184339.852302354 |
| 丰台区发生一起罐车泄漏事故 | 6月18日15时30分左右,在丰台区王佐镇南官459路公交场站内, | 6月18日15时30分 | 6/18/2015 15:30:00 | http://www.bjsafety.gov.cn/accidentin | 北京市丰台区 | 116.3184339.852302354 |
| 延庆县发生一起1人死亡事故 | 6月19日20时许,延庆县城东大街广配室内一人触电身亡,该 | 6月19日20时 | 6/19/2015 20:00:00 | http://www.bjsafety.gov.cn/accidentin | 北京市延庆县 | 116.0087540.458442001 |
| 房山区发生一起爆燃事故 | 6月14日9时30分许,房山区良乡拱辰大街9号“首品一家” | 6月14日9时30分 | 6/14/2015 09:30:00 | http://www.bjsafety.gov.cn/accidentin | 北京市房山区 | 116.1735539.732537984 |
| 石景山区发生一起3人受伤事故 | 31日19时40分,中铁建设集团工程资源分公司在石景山苹果园 | 31日19时40分 | 1/1/0001 00:00:00 | http://www.bisafetv.oov.cn/accidentin | 北京市石景山 | 116.1890039.923428953 |

disaster data crawling result

disaster data crawling result

Typical experiments

➤ Background map images search



How to automatically and accurately identify the map images about disaster?

Typical experiments

➤ Background map images search

Residential area map images around the site of explosion in Tianjin



交通部：芦山地震交通损失63亿 重建按三年

发稿时间：2013-07-03 10:23:38 来源：中国新闻网 中国青年

中新网7月3日电 据交通部网站消息，四川芦山地震灾区受损国道432公里、县

里、汽车

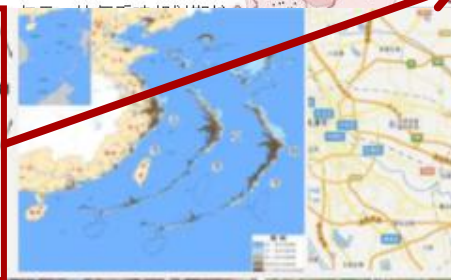
年7月至

消息

势，研究

门、各单

运行中在



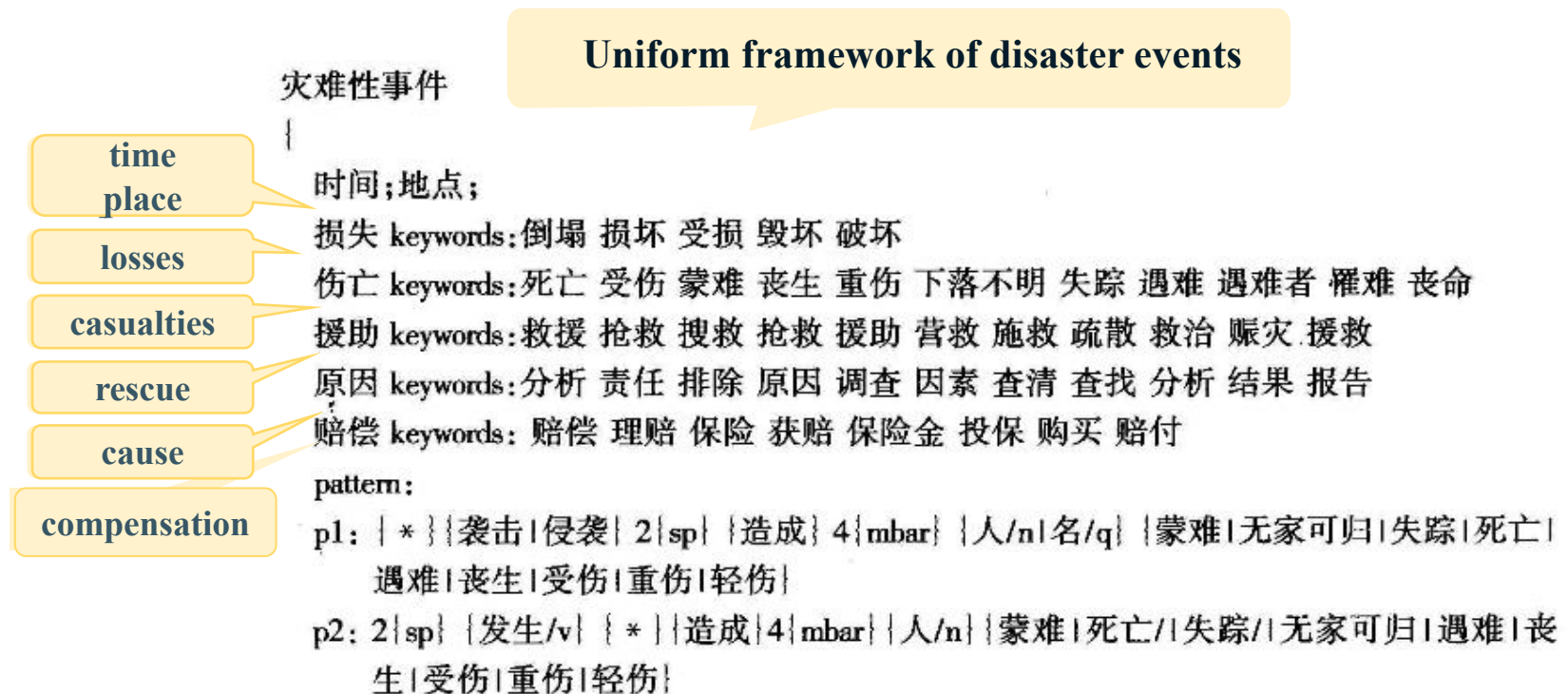
外，还巴占增设

Typical experiments

➤ Internet-based Disaster Information Extraction

Conditional Random Fields(CRF) was used to identify the event:

- Preprocessing: Word Segmentation
- Training for Information Extraction Model : used CRF++ tool



Typical experiments



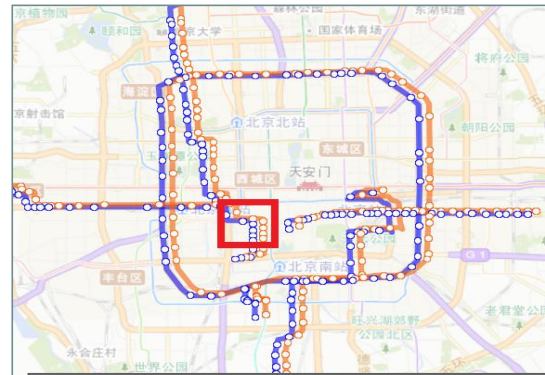
Typical experiments

➤ Integration of the Fundamental Geographical Information on the Internet

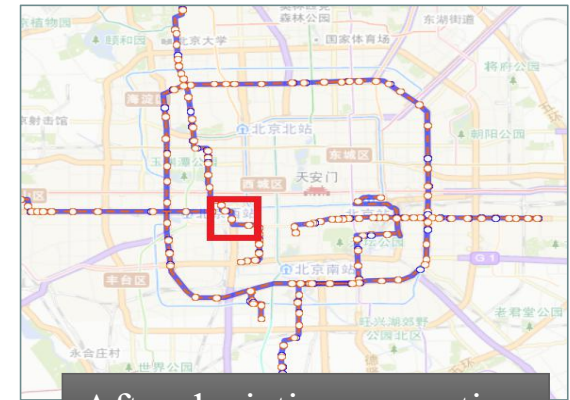
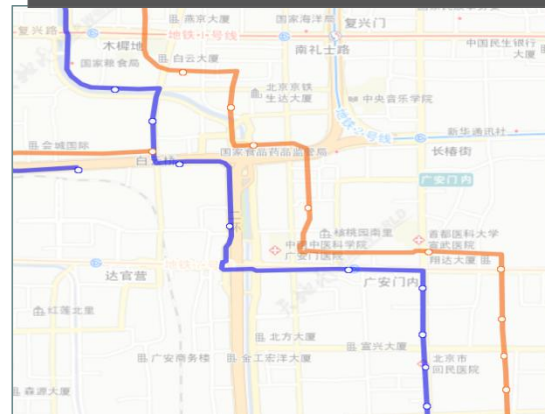
Deviation Correction:

Bus lines and POI data from different sources:

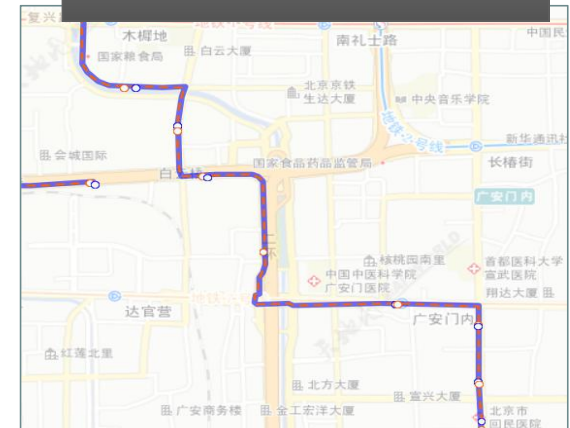
- serious spatial position deviation among the same points
- after correction, the deviation can be controlled within 10 meters



Before deviation correction



After deviation correction

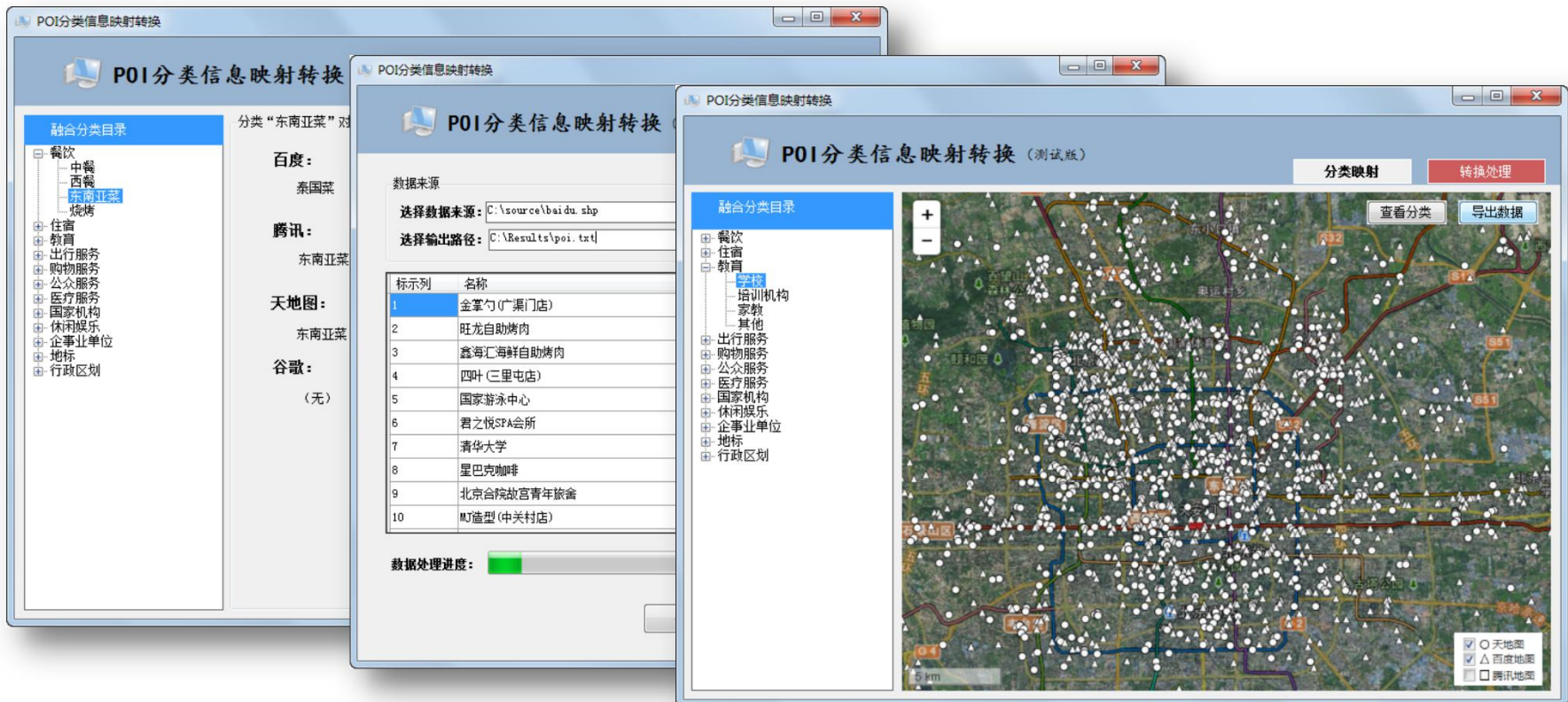


Typical experiments

➤ Integration of the Fundamental Geographical Information on the Internet

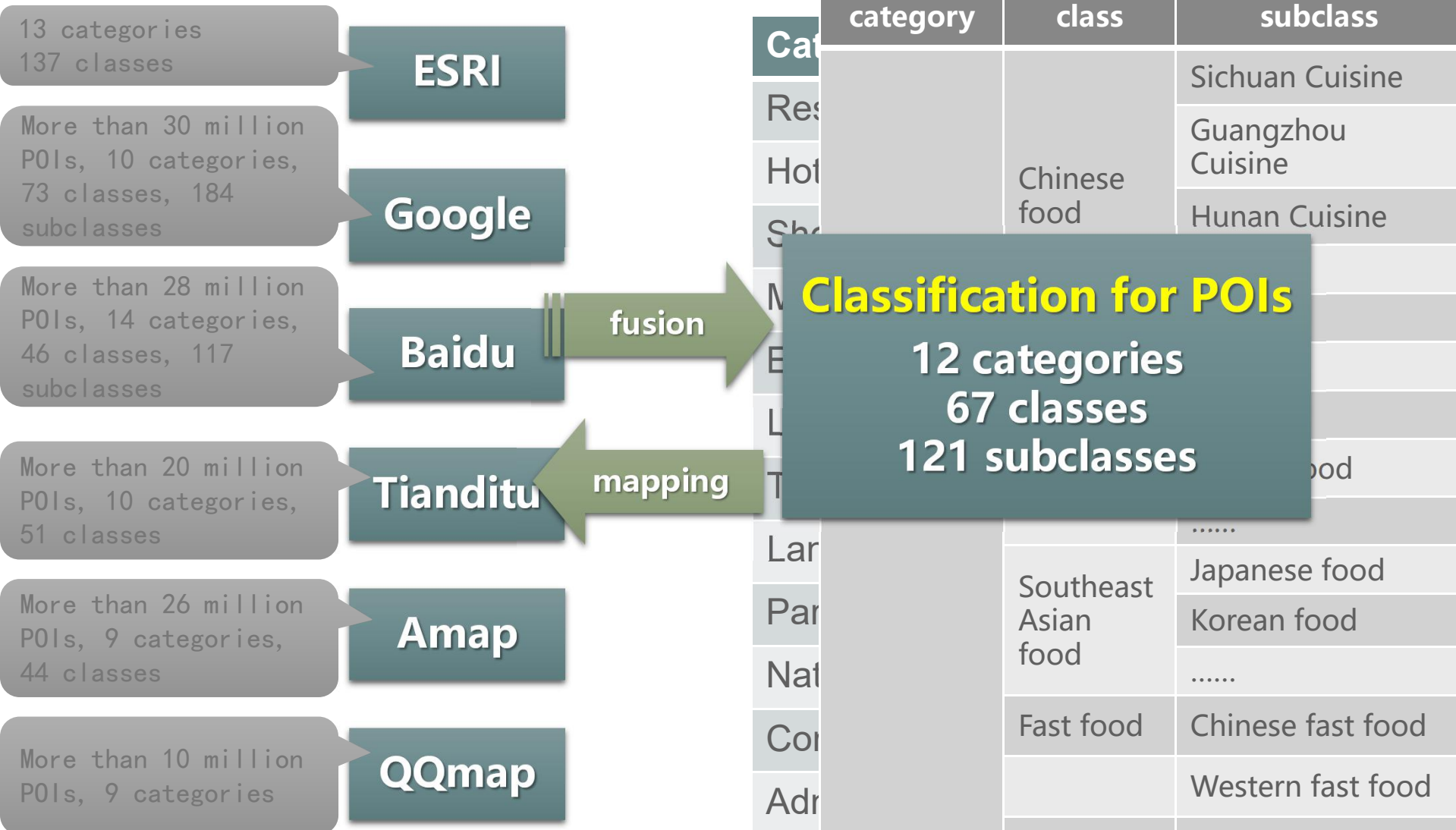
Multi-source POI classification information mapping and transformation:

- Multi-source POI classification information mapping
- Category attribute transformation processing
- Cross-site POI information extraction



Typical experiments

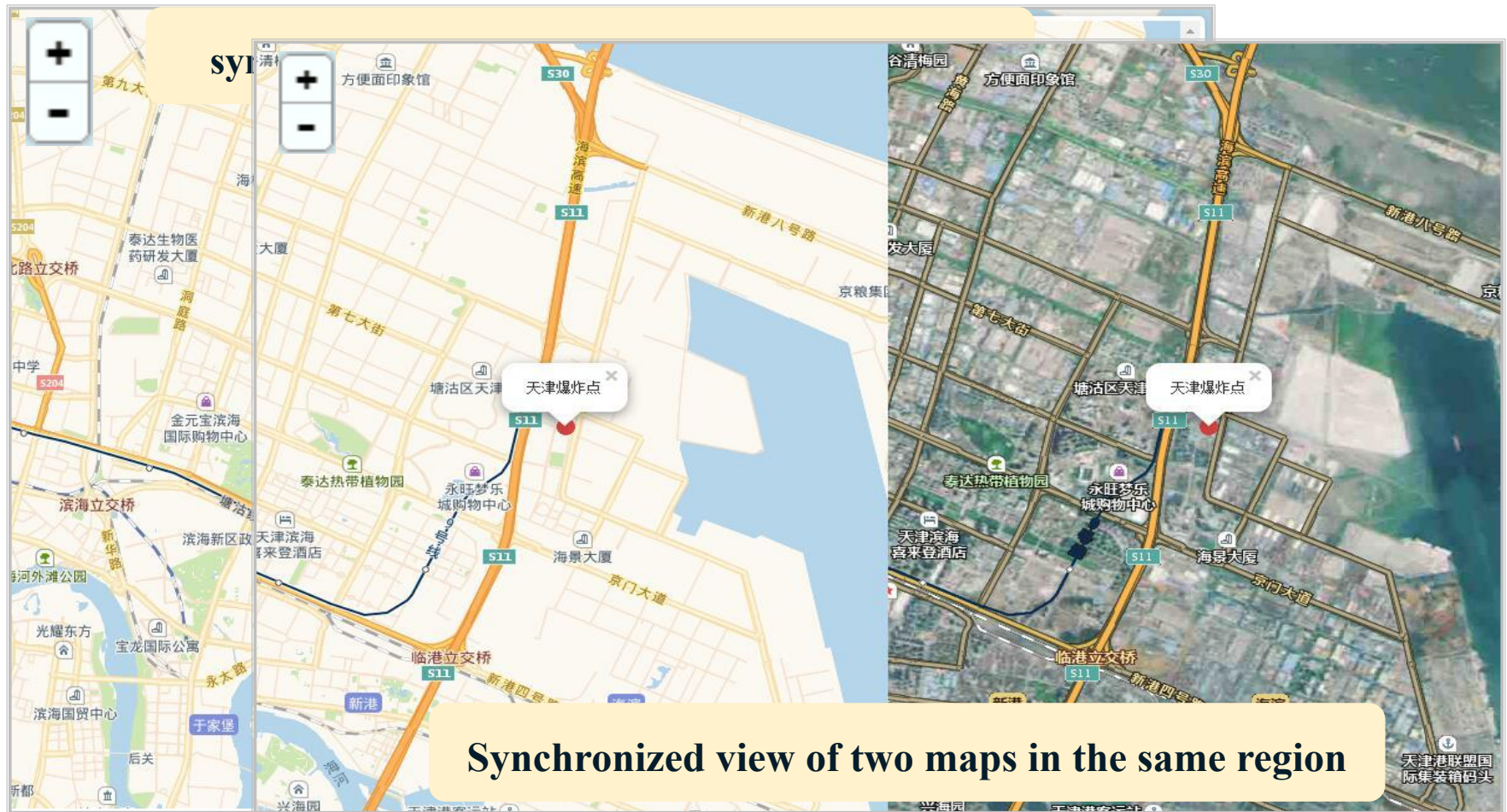
- Integration of the Fundamental Geographical Information on the Internet
Multi-source POI classification information mapping and transformation:



Typical experiments

➤ Integration of the Fundamental Geographical Information on the Internet

multi-source map data fusion:



Conclusion

1. Internet geographic information contains huge values for emergency response, damage evaluation and reconstruction.
2. For the massive, heterogeneous, unstructured internet geographical information, a radically new technical architecture would be adopted from the data acquiring, cleaning to data fusion and application.
3. Through several typical experiments, it is proved that it is feasible to discover and integrate disaster-related geospatial data from Web.

Event
Location

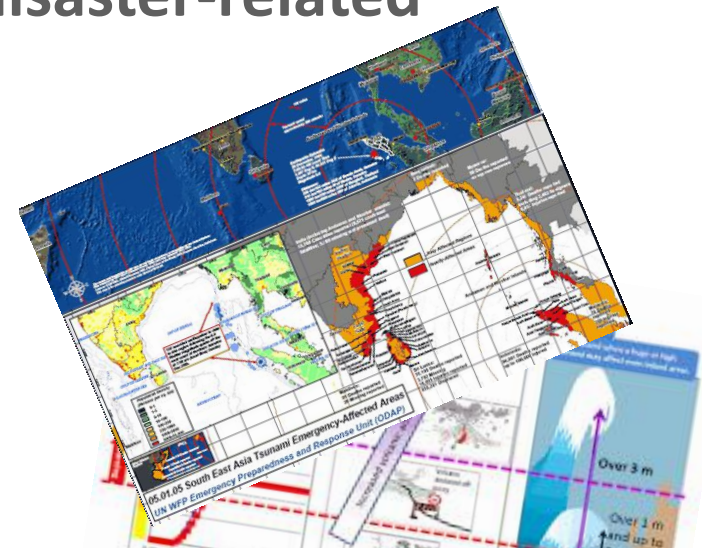
Evaluation Model

Disaster
Mapping

Disaster information semantic
framework

Semantic
extraction

Deep Web Searching



Ongoing work

1. Deep web searching technology

- It is necessary to develop high effective Internet search techniques to improve the speed of content traversal and information extraction.
- intellectualized, personalized, specialized and visualized search

2. Disaster information semantic framework

- A standard mapping framework for understanding and extracting the disaster information

3. Temporal and spatial position description and matching

- transform text description to geo-knowledge

4. Evaluation model

- Evaluate the reliability and availability of information sources, in order to develop the full value of Internet geographical information

5. Effective information integration framework



Thanks