Discuss on the Mode of Data Development in Internet Time

Wu Yuxian
Institute of Economics, CASS
The modes of data development:

*manual*: recording by tying, mailing by horse, write …..

*mechanical*: type, photography, microforms …..

*electronically*: recorder, digitized-meter, computer …..

*network*: www, web2.0, Grid, ….. GGG

*(Identity: collection, processing, transmission, storage, retrieval, use, etc*)
Two trends of network era

Trend in change of **technology**
(top-down)

www → Grid → …GGG

web2.0

(bottom-up)

Trend in change of **market needs**
Trend in change of market needs

Generations of web:

1.0  WWW (a loose title index)

1.5  portal engine, search engine
     (classific catalog, subject index)

2.0  BLOG, WIKI, IM, RSS, SNS
     (person to person, P2P)
Web1.0 ➔ 1.5: change for data development

1. Only to find data outer clues ➔ increasing orderliness and depth of information retrieval

2. Traditional heavy, backward production and processing methods ➔ a quasi-real-time, efficient method.

3. A complete chain and circuit of data provided, "self-learning mechanism" to be possible.

4. Web Data-Mining, a glimmer of developing data
Web2.0 : characteristics

1. The producer of data processing:
   websites → users → user → user.

2. Information services approach:
   “one-to-multi” → “multi-to-multi”

3. Information services functions:
   a simple "package" seek → provide in cluster
   (download, upload, publishing, storage)

4. Network efficiency: client-oriented web improve the channels and efficiency of data transmission greatly.

5. Network connectivity means:
   “web to web” → “person to person”, P2P
Trend in change of technology

Grid:
through standardized middleware platforms to organize mobilize resources, control data flow, balance load tasks, providing oriented-services to users, etc.

*For data development, the priorities of Grid:*

**Data Mining** technology:
the core of knowledge discovery.

**Data Warehouse** technology
Data Mining in the two trends

Web data mining:

Web2.0 focuses on: searching Information, filtering information, the quality of data.

Data mining for Data Grid:

Data Grid focus on: integration & modeling of data to support the complex data requires from the database perspective.

Function Merge:

Searching, filtering, integration, modeling
Trends: to be merged

Web 2.0:
- information resources from P2P terminal
- search engine
- web data-mining technology for IR

Data Grid:
- data-mining technology for DB
- self-learning mechanism
- data warehouse technology
New mode in the vision of NGN

Data Collection on P2P level + Data Mining:

Real-time data flow from terminal of P2P

Data Pool → Data Mining → Multi-stores Database

knowledge file

Databases

Data Warehouse
Implement scope

1. Enterprise:
data from client-server network + SNS (no server)

New mode

2. Internet
data from web2.0
co-operation & resource sharing

The third factor beside “technology” & “market”

Sharing = efficiency, quality
Cooperation = bridge

enterprise ↔ enterprise
Thanks