Bridging Different Generation of Web via Exploiting Semantic Social Web Blog Portal

Prof.(Dr.) Yuh-Jong Hu

June-28-2007

hu@cs.nccu.edu.tw
Emerging Network Technology(ENT) Lab.
Department of Computer Science
National Chengchi University, Taipei, Taiwan

1

Bridging Different Generation of Web

☞ *What do you mean bridging different generation of Web?*

    ✔ *To have data, information, knowledge sharing and integration*

    ✔ *To achieve services compatibility and interoperability*

    ✔ *To examine the architecture and characteristics features among different Web generation*

☞ *What are the incentives?*

    ✔ *To leverage the power of new generation Web technologies for old information.*

    ✔ *To provide the information sharing mechanisms between different Web generation.*

    ✔ *To populate the information from old Web generation to a new Web generation.*

Bridging Different Generation of Web (conti.)

☞ *Bridging and transition of different generation of Web is an evolution not a revolution.*

☞ *We must identify what are the primary differences among Web generations.*

☞ *We decided to exploit semantic social web blog as our first test plan.*

Bridging Different Generation of Web (conti.)

☞ *Web 1.0 (Static Web) ⇄ Web 2.0 (Social Web)*

    ✔ *Social Web Annotation*

    ✔ *Ongoing Process*

☞ *Web 1.0 ⇄ Web 3.0*

    ✔ *Semantic Annotation via web mining from resource's content*

    ✔ *Too Ambitious*

    ✔ *Automatic is almost impossible*

    ✔ *Only suitable for a narrow domain*

Bridging Different Generation of Web (conti.)

☞ *Web 2.0 (Social Web) ⇄ Web 3.0 (Semantic Web)*

    ✔ *Semantic Social Web Annotation*

    ✔ *Our Research Goal*

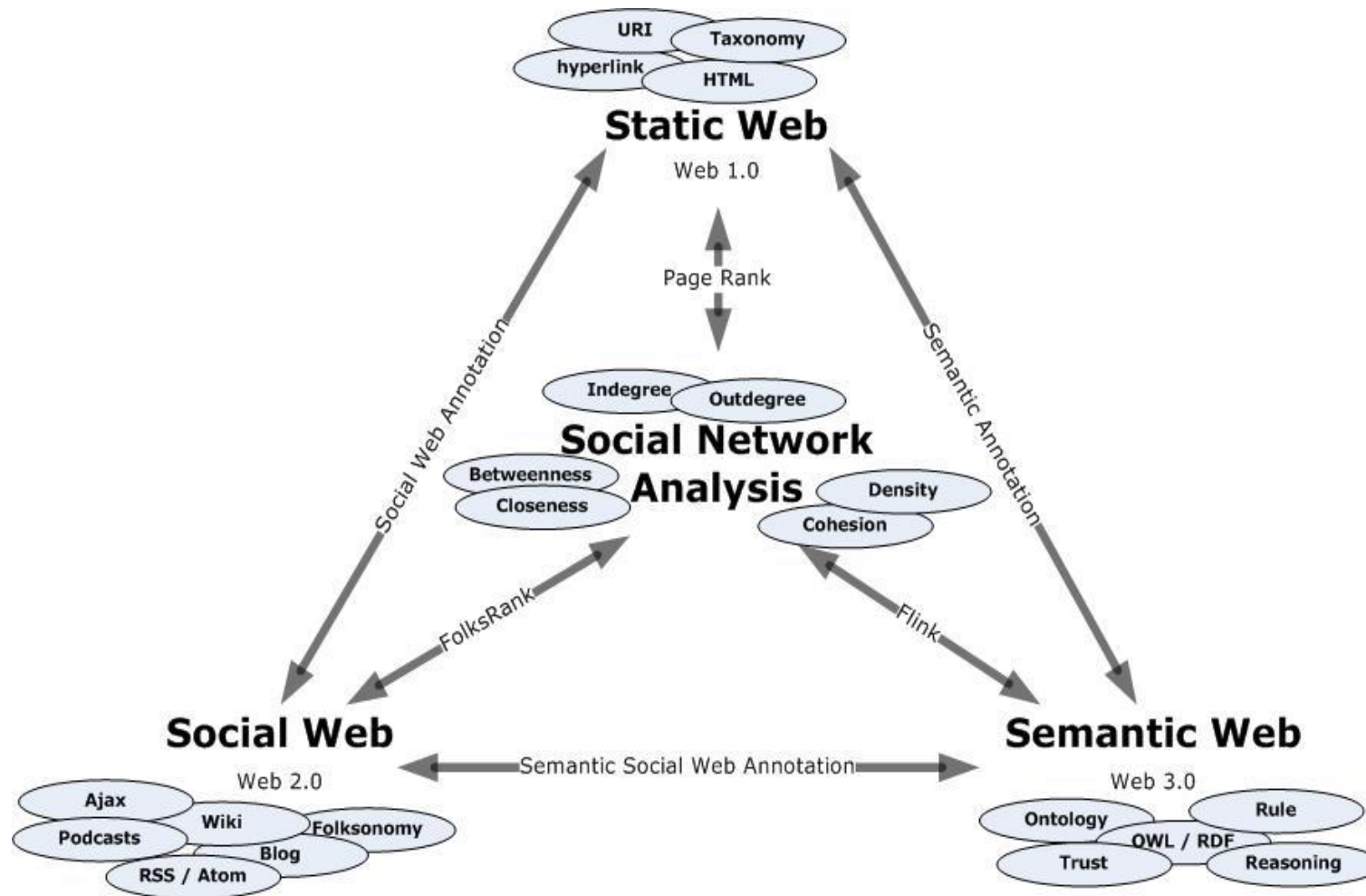   ✔ *We are dealing with user's tags instead of resource's content*

SNA ⇌ WWW (Web 1.0, Web 2.0, Web 3.0)

☞ *SNA ⇌ Web 1.0 (Static Web): Page Rank*

☞ *SNA ⇌ Web 2.0 (Social Web): FolksRank*

☞ *SNA ⇌ Web 3.0 (Semantic Web): Flink*

SNA ⇌ WWW (Web 1.0, Web 2.0, Web 3.0) (conti.)

What Are Blog's Problems?

☞ *Contents Are Distributed*

☞ *Formats Are Various*

☞ *Information Are Personal*

☞ *Information Are Voluminous*

✔ *Keyword-Based Search Is Not Enough*

✔ *Which Articles Are Important and Worthy To Be Read*

☞ *Information Syndication Is Weak*

Semantic Social Web Blog

☞ *Semantic Blog: Blojsom*

☞ *Tagging Blog Portal: Technorati, Blogpulse*

☞ *Semantic Social Web Blog Portal Elements:*

- ✔ *Targeted Blog Sites: WRETCH*

- ✔ *Crawler: From User → Friend → RSS → Article (Cache)*

- ✔ *Lightweight Ontology: Blog and Topic*

- ✔ *Social Network Analysis (SNA)*

- ✔ *Semantic Social Web Query Services*

Semantic Social Web Query Services (I)

☞ *I would like to search authors and their blog articles with "cuisine" tag paired with "restaurant" keyword in the associated title or content of the article collected from the entire blog community. Furthermore, please present these authors' names and their associated titles of article in a decreasing order of authors' indegree measures*

Semantic Social Web Query Services (I) (conti.)

```
prefix blog: <http://blog.nccucs.org/blog.owl#>
prefix rdf: <http://www.w3.org/19999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?Author ?Article
WHERE
{?Article rdf:type blog:Article
 ?Article blog:has_articleTag blog:cuisine
 ?Article blog:has_author ?Person
 ?Person blog:person_ID ?Author
 ?Person blog:person_indegree ?Popularity
 FILTER {regex(?TitleOfArticle, "restaurant") ||
 regex(?ContentOfArticle, "restaurant"))}
   }
ORDER BY DESC (?Popularity)
```

Semantic Social Web Query Services (II)

☞ *I would like to search blogger names and their articles from the cuisine channel for those of whom are known by authors presented in scenario one. Furthermore, please present these blogger names and their associated titles of article in a decreasing order of authors' indegree measures.*

Semantic Social Web Query Services (II) (conti.)

```
prefix blog: <http://blog.nccucs.org/blog.owl#>
prefix rdf: <http://www.w3.org/19999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?Author ?Friend ?TitleOfFriendArticle
WHERE
{........
 Codes Same As Scenario One
 ....
?Person blog:has_knows ?friend
?friend blog:person_ID ?Friend
?FriendArticle blog:has_author ?friend
?FriendArticle blog:has_channel blog:CuisineChannel
?FriendArticle blog:article_title ?TitleOfFriendArticle
?FriendArticle blog:article_description ?ContentOfFriendArticle
FILTER {regex(?TitleOfFriendArticle, "restaurant") ||
regex(?ContentOfFriendArticle, "restaurant"))}
  }
ORDER BY DESC (?Popularity)
```
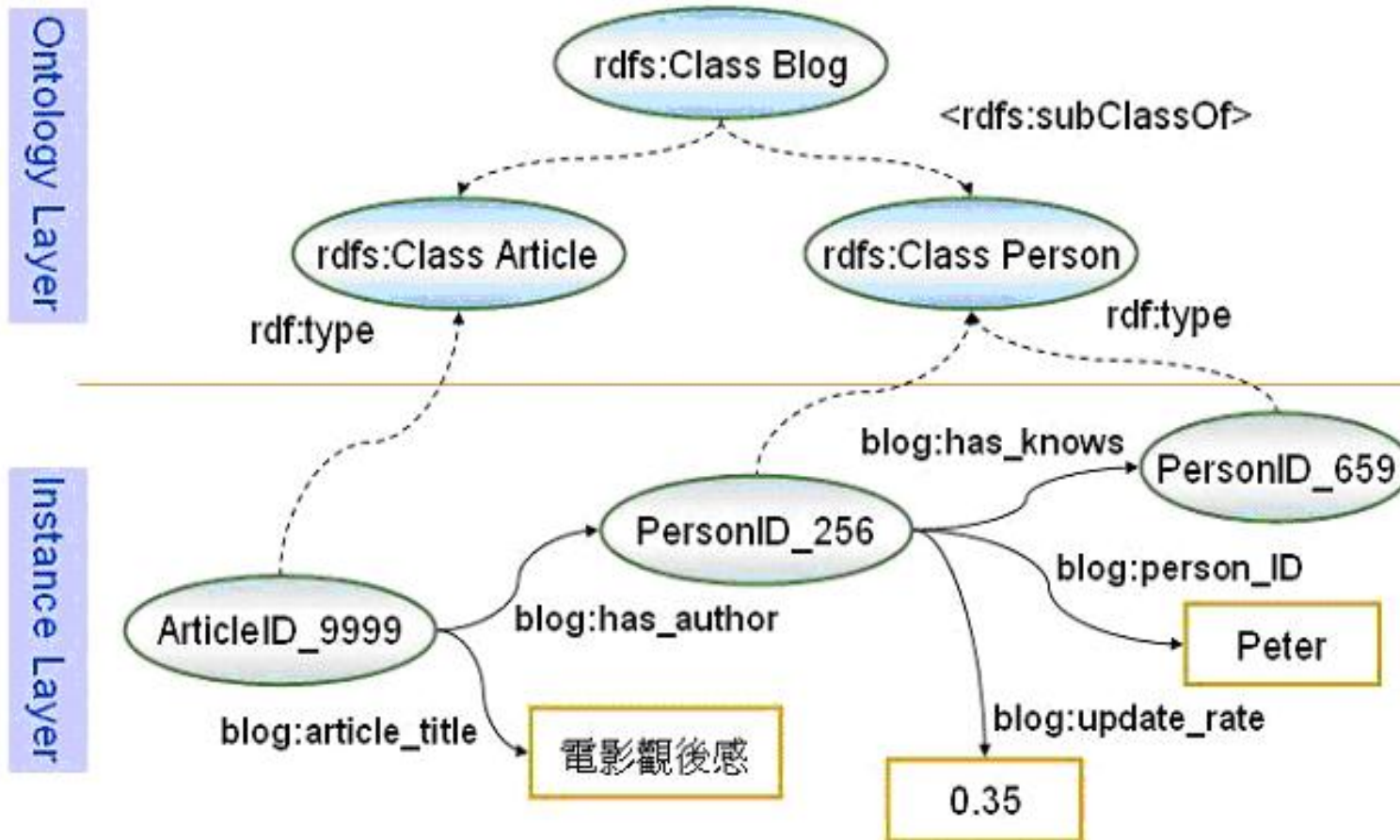
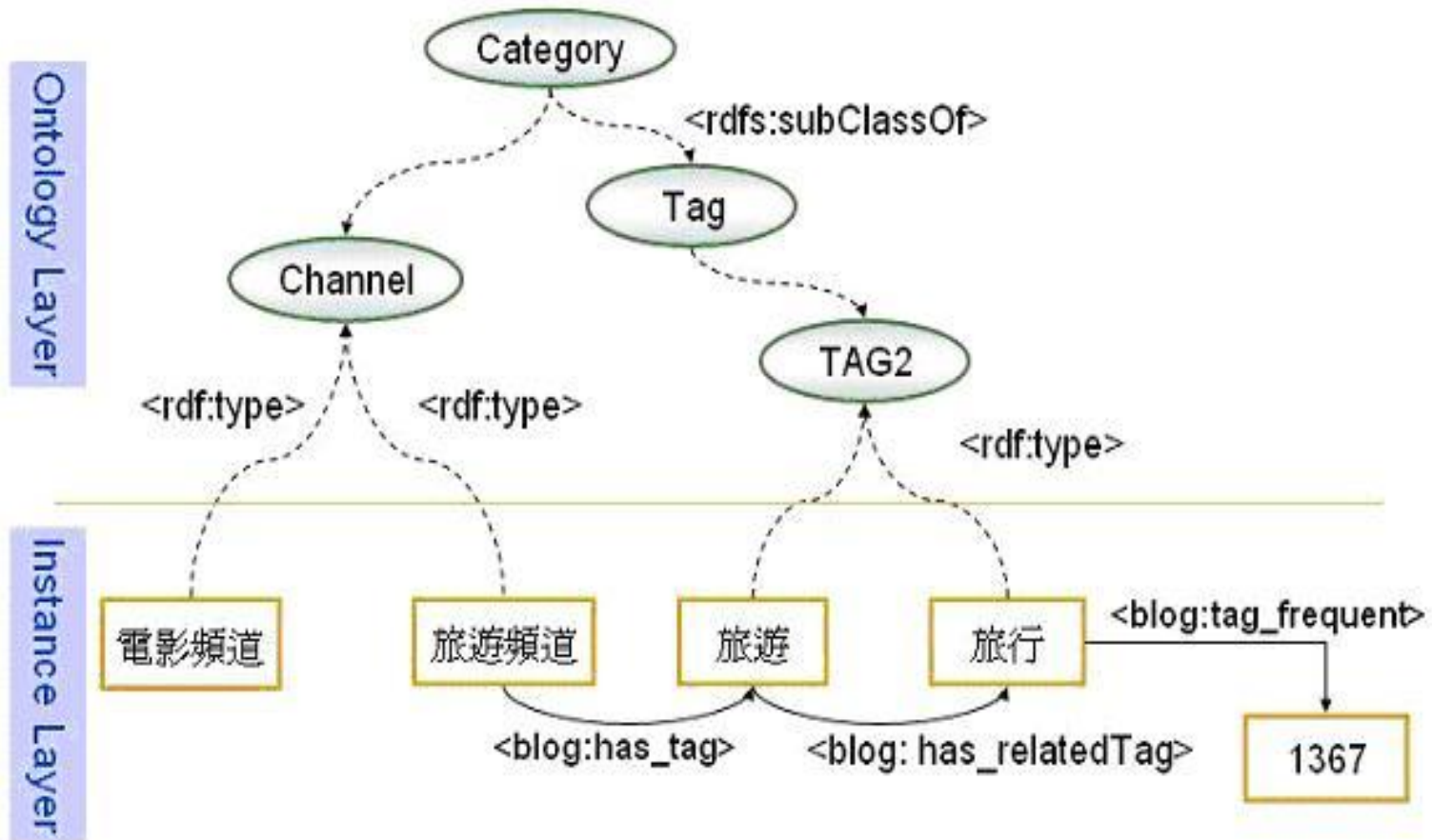# *Conceptual Schema for Semantic Social Web Blog Portal*
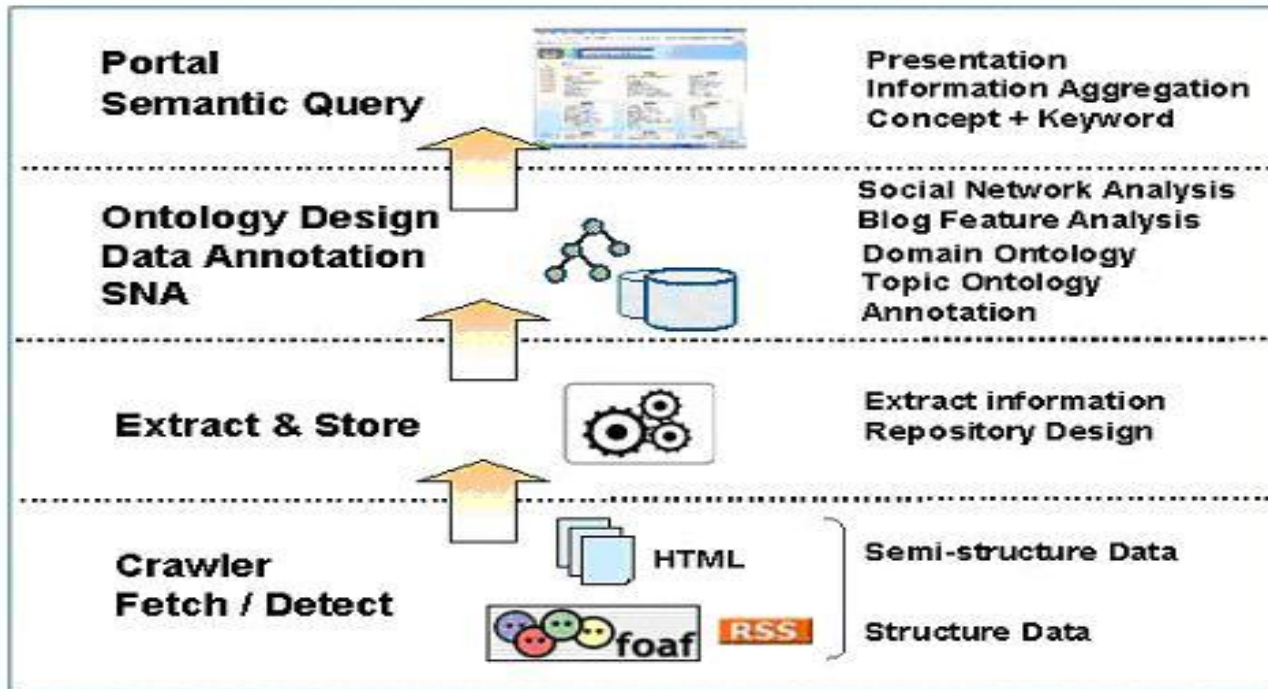
# The Blog Ontology
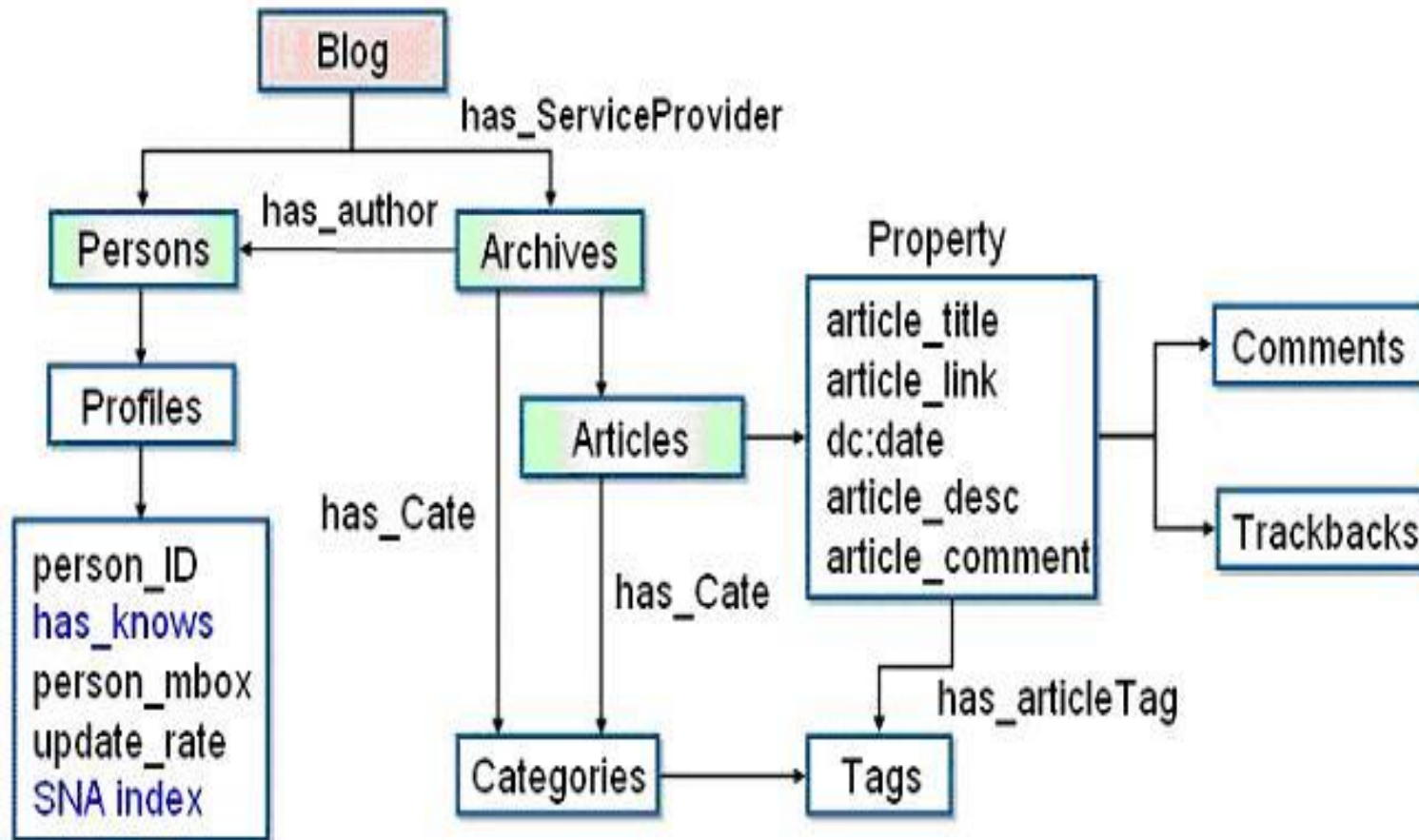
# The Blog Ontology for Article and Blogger

# *The Topic Ontology*

# Blog Social Network Status for WRETCH

☞ *Users Number: 960,771 → 140,000 → 108,518*

☞ *Number of Relationships: 1,296,986*

☞ *Entire Network Density Is Pretty Sparse: 0.011%*

☞ *Pajek SNA Software for*

  ✔ *Individual Level: Centrality*

  ✔ *Dyadic Level: Distance, Reachability*

  ✔ *Subset Level: Cohesive Subgroups*

  ✔ *Network Level: Density, Diameter, Centralization*

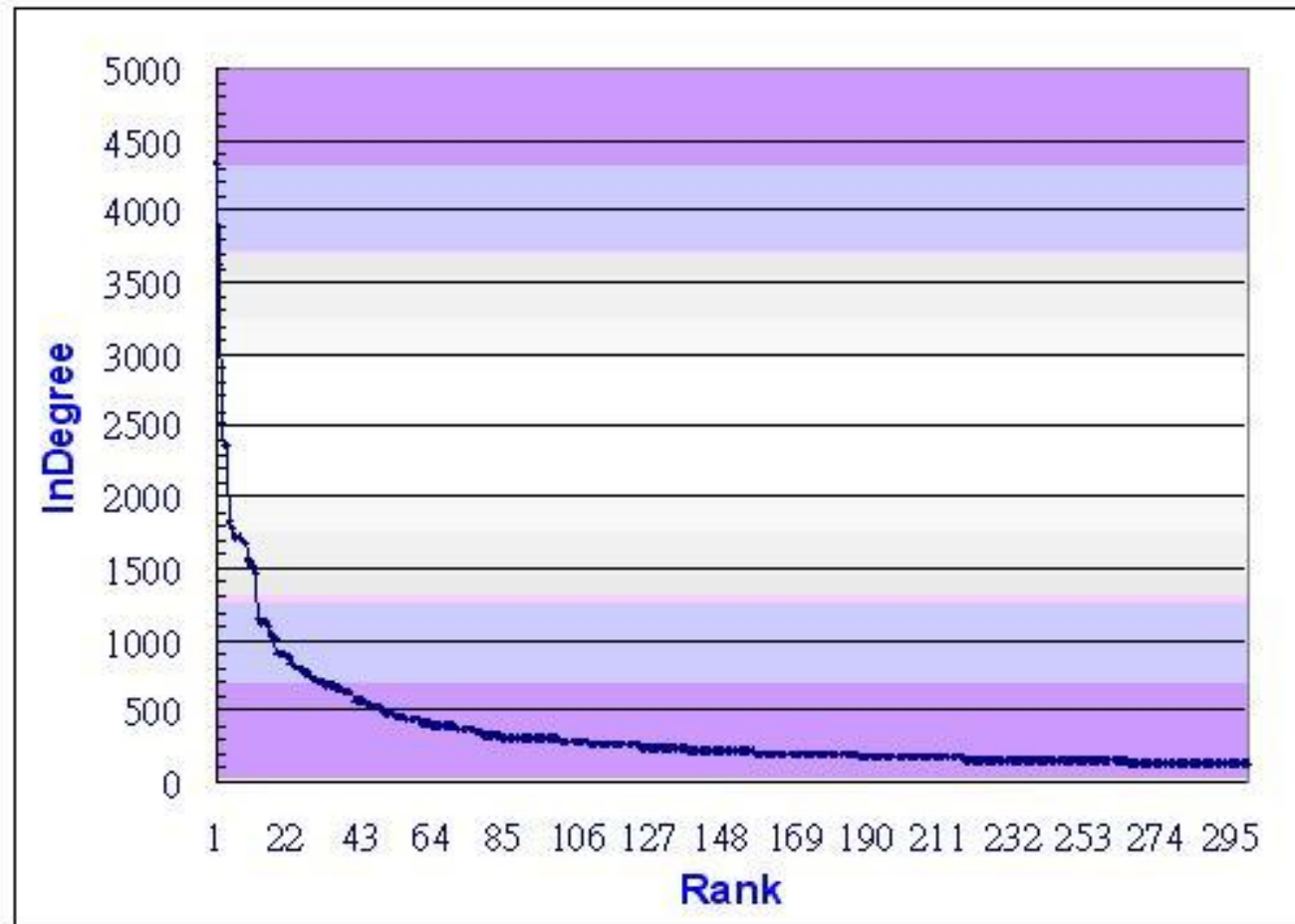# SNA Individual Level for Centrality

☛ *InDegree: Popularity, Prestige*

☛ *OutDegree: Category*

☛ *Closeness*

☛ *Betweenness*

# SNA Measures:
## Power Law Distribution of InDegree for Tags

# SNA Measures: Indegree Table

| Top 20 of InDegree | | | | | Top 20 of OutDegree | | | |
|---|---|---|---|---|---|---|---|---|
| Rank | Vertex | InDegree | Id | | Rank | Vertex | OutDegree | Id |
| 1 | 19 | 2693 | cwwany | | 1 | 13909 | 283 | super101dog |
| 2 | 297 | 1817 | gyapower | | 2 | 2214 | 272 | pink1640 |
| 3 | 115 | 1747 | SHE911 | | 3 | 965 | 247 | sushi88 |
| 4 | 282 | 1332 | J90909490 | | 4 | 4309 | 245 | lovedear |
| 5 | 108 | 1286 | fanwenshan | | 5 | 2288 | 154 | ceramics0811 |
| 6 | 110 | 1281 | jjboy | | 6 | 1005 | 152 | alice0103 |
| 7 | 327 | 1275 | wia627 | | 7 | 62280 | 147 | dennis0910 |
| 8 | 107 | 1254 | eddie0324 | | 8 | 3266 | 147 | wzj1206 |
| 9 | 277 | 1075 | a80906 | | 9 | 2612 | 145 | vanessa00 |
| 10 | 4112 | 897 | leigo | | 10 | 2966 | 144 | karen80577 |
| 11 | 2774 | 849 | mamiqqq | | 11 | 3939 | 141 | mazdoor |
| 12 | 112 | 822 | nicholas01 | | 12 | 316 | 141 | amile |
| 13 | 2166 | 818 | angela5247 | | 13 | 334 | 139 | apeiian |
| 14 | 2096 | 745 | arronbubest | | 14 | 27860 | 138 | andyfishwang |
| 15 | 117 | 656 | TANKmusic | | 15 | 4298 | 138 | alion |
| 16 | 2601 | 653 | bboy1030 | | 16 | 3358 | 135 | PMANA |
| 17 | 7338 | 644 | nolava | | 17 | 3355 | 135 | iomyself |
| 18 | 118 | 634 | yvonne1116 | | 18 | 325 | 129 | hayado |
| 19 | 3190 | 620 | Giddens | | 19 | 2331 | 129 | b89508152007 |
| 20 | 2255 | 617 | aprilkwc | | 20 | 3469 | 128 | BOON |

# SNA Measures: Clonseness and Betweenness

| | Top 20 of Closeness | | | | Top 20 of Betweeness | | |
|---|---|---|---|---|---|---|---|
| Rank | Vertex | Value | Id | Rank | Vertex | Value | Id |
| 1 | 19 | 0.3114890 | cwwany | 1 | 327 | 0.0137165 | wia627 |
| 2 | 297 | 0.2997216 | gyapower | 2 | 191 | 0.0114645 | ilove7388 |
| 3 | 115 | 0.2942667 | SHE911 | 3 | 297 | 0.0094941 | gyapower |
| 4 | 327 | 0.2932771 | wia627 | 4 | 2774 | 0.0067478 | mamiqqq |
| 5 | 108 | 0.2906045 | fanwenshan | 5 | 965 | 0.0060157 | sushi88 |
| 6 | 107 | 0.2895195 | eddie0324 | 6 | 2601 | 0.0058224 | bboy1030 |
| 7 | 282 | 0.2865868 | J90909490 | 7 | 13909 | 0.0056047 | super101dog |
| 8 | 110 | 0.2820530 | jjboy | 8 | 193 | 0.0054250 | minimamie |
| 9 | 2774 | 0.2815346 | mamiqqq | 9 | 531 | 0.0052539 | ppkeigo |
| 10 | 277 | 0.2764172 | a80906 | 10 | 429 | 0.0052008 | fjumonkey |
| 11 | 191 | 0.2750786 | ilove7388 | 11 | 3745 | 0.0045046 | kerina |
| 12 | 1761 | 0.2748988 | beahayden | 12 | 730 | 0.0044758 | sasakinki |
| 13 | 2166 | 0.2725833 | angela5247 | 13 | 4309 | 0.0039384 | lovedear |
| 14 | 4112 | 0.2701986 | leigo | 14 | 4061 | 0.0036804 | redduck |
| 15 | 118 | 0.2687906 | yvonnel116 | 15 | 3190 | 0.0036214 | Giddens |
| 16 | 339 | 0.2687765 | princezzz | 16 | 7271 | 0.0035771 | peggyegg |
| 17 | 2096 | 0.2670085 | arronbubest | 17 | 330 | 0.0035287 | kamel988428 |
| 18 | 742 | 0.2656541 | vul3e94 | 18 | 2612 | 0.0031776 | vanessa00 |
| 19 | 117 | 0.2653311 | TANKmusic | 19 | 2502 | 0.0030712 | imjaly2 |
| 20 | 2707 | 0.2652314 | IMXIAOXIAO | 20 | 7152 | 0.0030282 | lisahiphop |

# *SNA Subset Level for Cohesive*

☞ *Strong/Weak Components*

☞ *K-Cores: 5-Cores, 9-Cores*

☞ *Cliques*
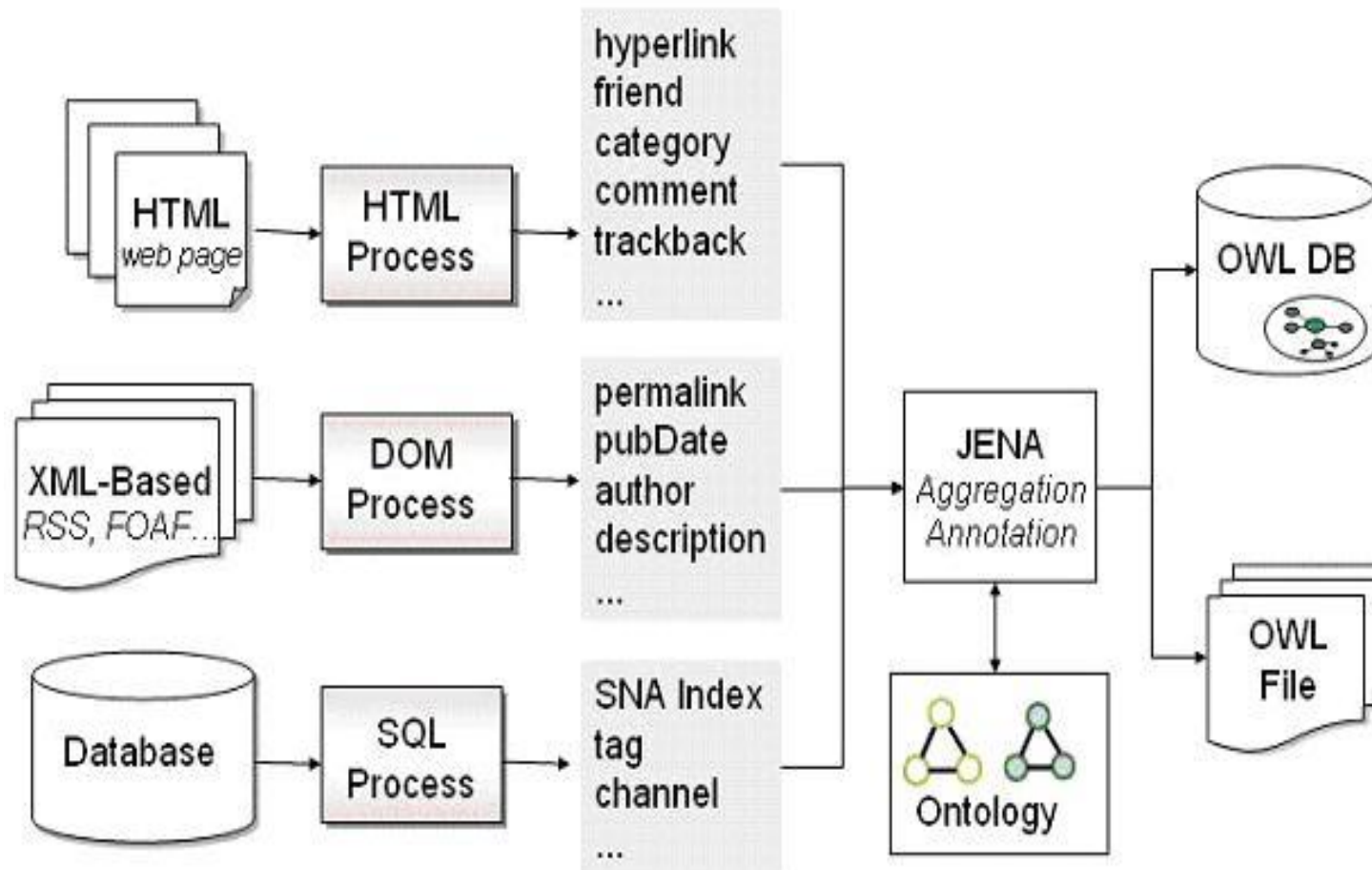
# SNA Measures: 5-Cores and 9-Cores
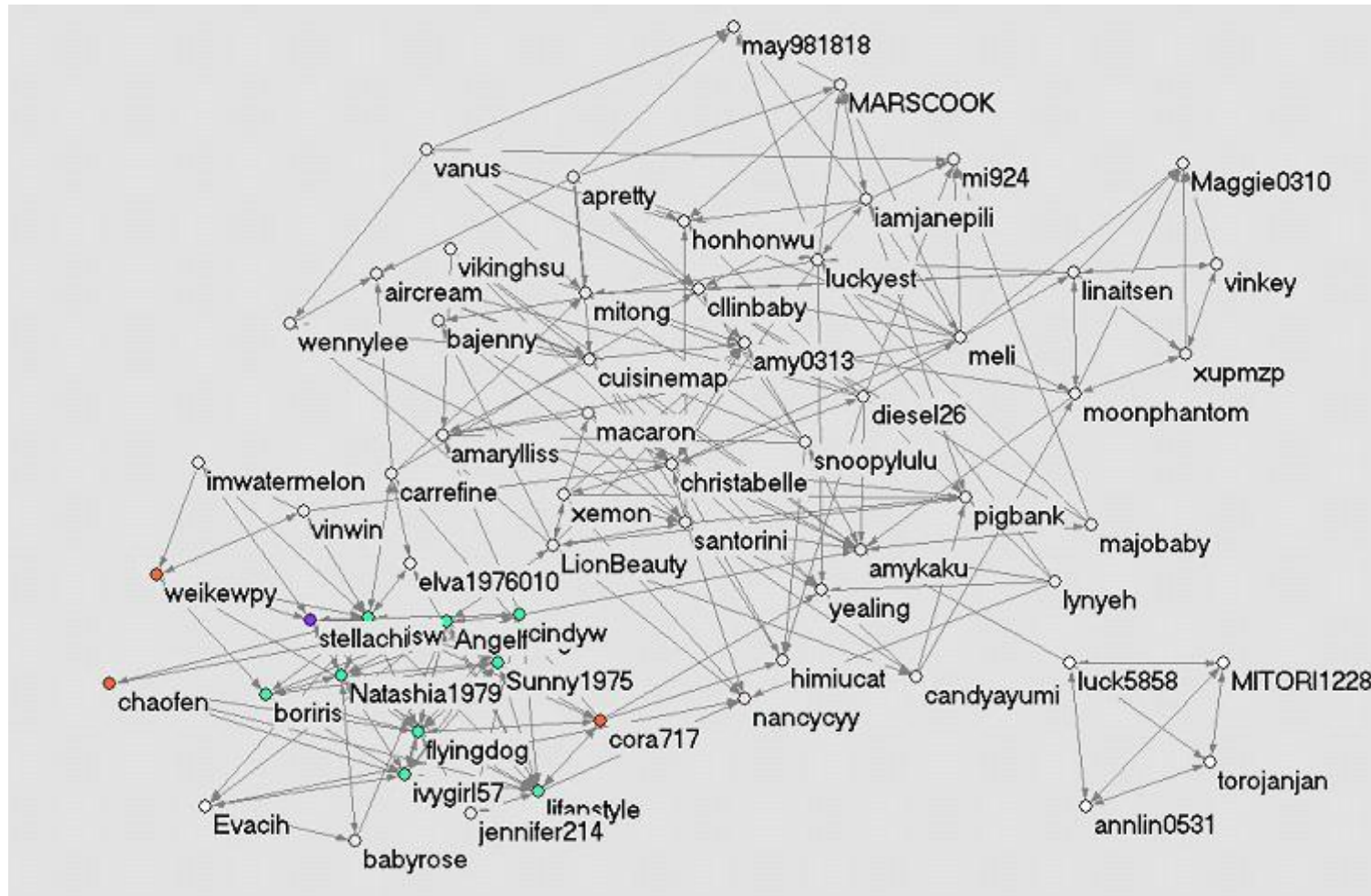
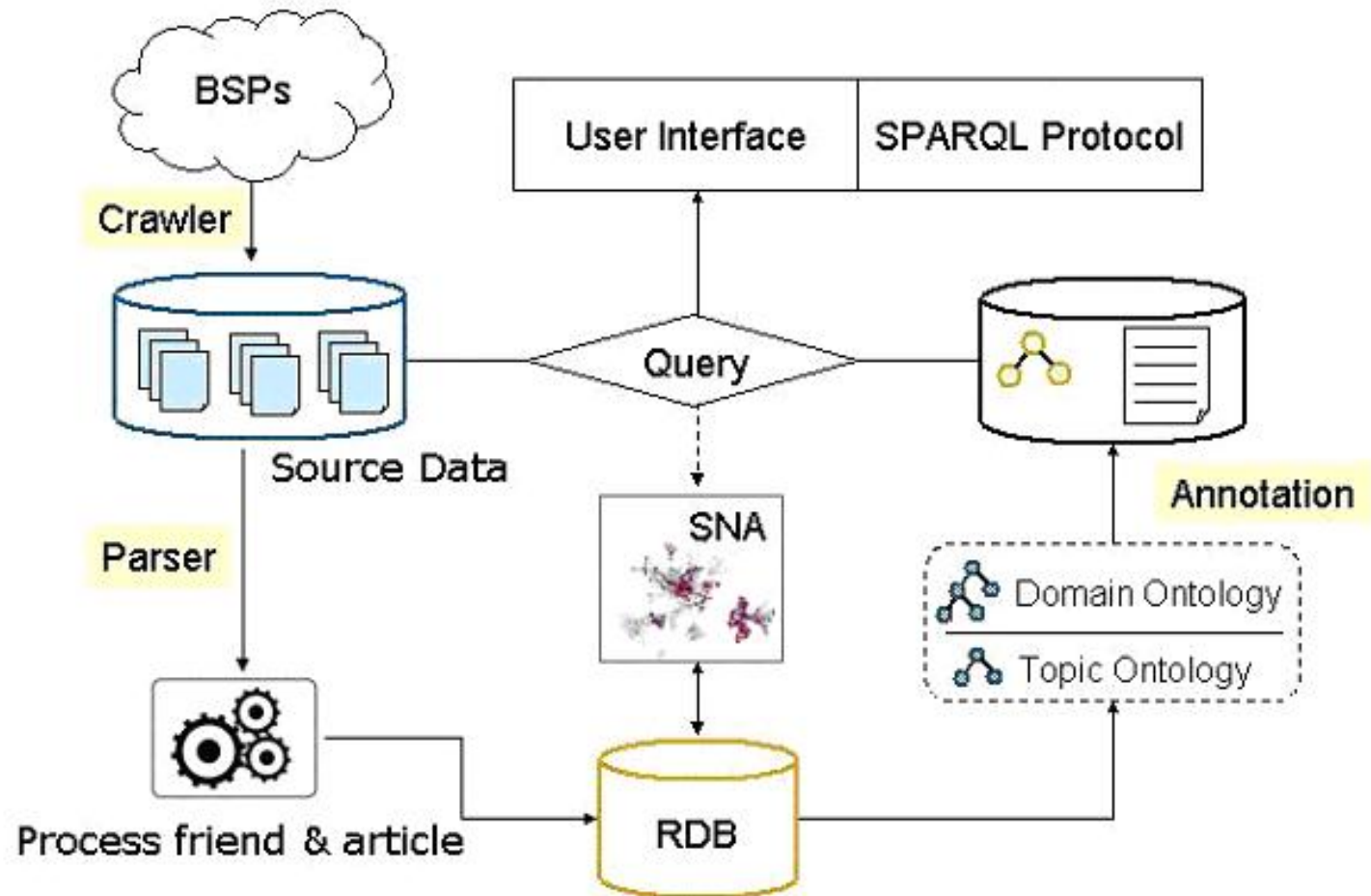# *Implications for SNA ⇄ WWW (Web 1.0, Web 2.0, Web 3.0)*

☞ *SNA ⇄ Social Web*

☞ *SNA ⇄ Semantic Web*

☞ *Social Web ⇄ Semantic Web*

# Semantic Social Web Annotation

# A Framework for Semantic Social Web Blog Portal

# *Conclusion*

☞ *Social web consensus can be derived from tags cloud convergence but whether $\xrightarrow{?}$ social intelligence.*

☞ *Bridging Web 2.0 $\rightleftarrows$ Web 3.0 via folksonomy + ontology is a trivial approach, e.g.*

☞ *Folksonomy+Ontology $\rightarrow$ semantic social web annotation (SSWA).*

☞ *We are not sure whether folksonomy alone can be demonstrated as SSWA, i.e., folksonomy $\xrightarrow{?}$ SSWA.*

# Conclusion (conti.)

☞ *We are working on tagging system with automatic generation of se-mantic hierarchical relationships among tags.*

☞ *SNA-based semantic social web query services can be achieved for some blog community characteristics discovery.*

☞ *Can we generalize our approach for other Web 2.0 domains, such as Wiki, Web Services, etc.*

☞ *What are the implications for:*
*social web $\rightleftarrows$ SNA $\rightleftarrows$ semantic web $\rightleftarrows$ social web?*