Toward Scholarly Event Digital Library Services

2008.06.16

Senator Jeong

http://bike.snu.ac.kr
BioMedical Knowledge Engineering Lab., Seoul National University
Contents

1. Introduction
2. Current Work
3. Future Work Plan
4. Research Implications
Research Tracks

Track 1
- Scholarly Event Ontology
- Data Harvesting

✔ Backbone to Describe, collect, organize, mine, analyze Scholarly Event Data

Track 2
- Mining scholarly meaningful information

✔ Model, Scholar & Event Evaluation metrics
✔ Determine Prominence & Quality

Track 3
- Social Network based info. platform

✔ SNA of Event Related Data
✔ KOS Generation
✔ Research Trends
Contents

1. Introduction
   - Scholarly Event
   - Scholarly Information Needs
   - Research Goal & Action Plan

2. Current Work

3. Future Work Plan

4. Research Implications
1.1. Scholarly Event

• Scholarly Event: **Conferences, Workshops, Seminars, Symposia**

• A sequentially and spatially organized collection of scholars’ interactions with the intention of delivering and sharing knowledge, exchanging research ideas, and performing related activities.

• Some of the major scholarly communication channels

• They provide scholars with opportunities to
  – Publish up-to-date scientific research results,
  – Get feedback from scientific communities, and
  – Exchange research interests and ideas with each other
  – Demonstrate current research trends

• Valuable information sources
1.2. Scholarly Information Needs

- **Information Need 1**: Information of a simple nature
- **Information Need 2**: In-depth Information for scholarly meaningful analysis
- **Information Need 3**: Subject Context Information
1.2.1. Information of a simple nature

- **Which conferences** will be (or have been) held on **which topics** and
- **When, where** they are held and **who will organize** them, and
- **What the due dates are** for submissions.
- Though this is basic information, it is **not often the case** that a scholar **gets a full and exhaustive picture** of scholarly events held in the world.
- Due to the **sheer volume** of scholarly events held by **various academic societies** and organizations
- **No one specific information channel** has been successful at keeping track of ever-growing conferences and providing their information to scholars.
1.2.2. In-depth Information Need

- **Who is regarded as prominent in a specific topic;**
- **What scholars are best suited for questions, consultations, and collaboration;**
- **Which conference is the most prominent in a field of study.**
- Researchers (newcomers in a given field) and research performance evaluators may have special interests.
- These information needs might be met partially at a minimal level since almost all conference websites list event general chairs, organizing committee and program committee members, invited speakers and/or award winners.
- However, since meaningful information regarding scholarly events is not provided by existing libraries in a sufficient and integrated manner, searching for relevant information can be a cumbersome endeavor.
1.2.3. Subject context information

- **How the research trends are changing?**
- **Which research topics are emerging or obsolete?**
- **Who (which conference) is related to whom in which context (topics)?**
- **Which conference is closest to my latest research?**

- **Existing information systems do not illuminate these traits sufficiently.**
1.3. Research Goal & Action Plan

• Assumption:

• *If researchers are able to get more comprehensive, meaningful, and context-sensitive information regarding scholarly events, then better scholarly communication will result.*

• Project’s objective is to satisfy scholars’ information needs
1.3. Research Goal & Action Plan

1. To build artifacts for satisfying scholars’ basic information needs by collecting and archiving scholarly event information.
   - **Prerequisite:**
     - To define an Scholarly Event Ontology
     - To build a system to harvest event data in a (semi)-automatic way.
1.3. Research Goal & Action Plan

2. To meet scholarly meaningful information needs by

- Mining scholarly meaningful information from harvested data.
- Designing a model and system to measure and determine scholars’ performance and the qualities of events.
1.3. Research Goal & Action Plan

3. To provide a platform to satisfy subject-context information needs.
   – Push for an automatic system to analyze and synthesize scholarly event data using social network metrics with special attention to subject context.
Contents

1. Introduction

2. Current Work
   - Scholarly Event Ontology Design
   - Evaluation Metrics for Scholars & Scholarly Events

3. Future Work Plan

4. Research Implications
2.1. SEDE: Ontology

• A partial response to simple information needs

• To develop a scholarly event ontology
  – with logical construction across multiple domains in detail,
  – capturing as many different event types as possible

• SEDE

• Scholarly Event DEscription Ontology
2.1. SEDE: Purpose

- To provide a *backbone* for:
  1. Describing, collecting and sharing scholarly event data and information
  2. Allowing event organizers and libraries to organize and manage event information for alerting relevant users more easily
  3. Evaluating the quality of researchers and events, and determining their prominence index through evaluation metrics
  4. Enabling software agents to extract and mine scholarly meaningful information from events
  5. Analyzing social relationships between events, subjects, organizations, and participants within/across domains
2.1. SEDE: Use-Case Scenarios

1. Event Data Harvesting
   - Event Website
   - Web Crawler
   - OntoCA (Annotation Tool)
   - Crawl Event URL
   - Annotate data using SEDe & OntoCA

2. Event Data Archiving & Alert Service
   - Scholarly Event Digital Library
   - SDI service using RSS
   - User
   - Search & retrieve events

3. Event Information Publishing
   - SEDe-embedded Editor
   - Edit and publish event information

4. Event Topic based Ontology Learning
   - SEDe instance data
   - Extract topics
   - Generated Ontologies

5. Evaluate Researchers or Events
   - Calculate and determine Researchers or Events’ Ranking
   - Rule:
     - If an event has many participants who have highly cited papers
     - Then it will be ranked high.

6. Event Similarity Calculation
   - Measure distance and similarity among events using event topics

JCDL2008
Doctoral Consortium
2.1. SEDE: Requirements

SEDE should be

1. Both in human and machine-readable format (RDF/OWL)
2. Interchangeable and evolvable
3. Scalable to describe events
4. Multi-functional
2.1. SEDE: Design Principles

1. Whenever possible, utilize existing metadata and ontologies rather than Re-invent existing wheels
   - SWRC for research
   - FOAF for people
   - SKOS for event topics
   - Dublin Core for metadata

2. From a survey and analysis of various event types such as conferences, workshops, symposia, seminars, and webinars as many scholarly event-specific properties and attributes as possible are being extracted.
2.1. SEDE: Modeling: References

- SKOS Topic
- FRBR Publication
- SWRC Research Community
- DublinCore metadata
- FOAF Person/Group
2.1. SEDE: Issue

- Although the design of SEDE aims to be uncontroversial and consistent with the generally accepted view of scholarly events,
- it is inevitable that some of the design decisions may be debatable.
- Therefore SEDE will be open for anyone to contribute to its improvement and modification via the thesis project wiki

- www.eventography.org
2.1. SEDE: Existing Ontologies

- ESWC2006 Conference ontology
  - it only fits the case of the ESWC2006 conference because no event information has exactly the same structure and terminology.
    - a session is grouped around a specific time interval,
    - A session can mean a track grouped around a specific research topic or theme.
  - it is designed at a level of description that does not support common grounds for semantic interoperability.
  - SEDE attempts to build a more axiomatic scholarly event description approach
2.1. SEDE: Existing Ontologies

- The Iugo project extends the ESWC2006 conference ontology.
- It is designed to present data for discovery and retrieval of materials that relate to particular events at the user interface level.
- It inherits ESWC2006 ontology structure.
  - The model is not sufficient to be understood by entities that publish and harvest scholarly event data, and infer scholarly meaningful information from them.
- Scalability appears not to have been a major concern when designing the Iugo project.
2.1. SEDE: Existing Ontologies

- ITTALKS was built to provide access to information technology related talk announcements.
- Its scenario is to provide the inferred information that corresponds to the user’s profile, personal schedule, location at the time of the event, distance, and usage patterns by integrating service agents with online information resources.
- It may be effective in a small domain, but is not generally applicable to all types of scholarly event and domains.

- SEDE covers not only event announcements but also all related entities to infer scholarly meaningful information.
2.1. SEDE: What Stage is it in?
2.2. Evaluation Metrics

- Responding to 2\textsuperscript{nd} Information Needs,
- A New and evaluation metrics for scholars’ performance & scholarly events are modeled.
- In searching for an alternative evaluation method (not compete with other methods)
- Scholars and scholarly events’ Prominence are to be determined based on the data indexed in SEDE Knowledge Base.
2.2. Evaluation Metrics: For Scholar

- Citation-based measures
  - shortage in coverage, incorrectness of data in citation index databases, self-citations, and even negative citations.
  - Alter or make them more complete; h-index, g-index, PageRank, Usage Impact Factor, and so on.

- Usage Impact Factor
  - evaluates referenced work by exploiting usage (download) frequency in the bibliographic repository.
  - Retrieval Coherence Assumption based
  - Use social network metrics to determine the impact of the particular work on other works.
  - It is arguable because there is no way to see if the downloaded work was really used to write one’s work.
2.2. Evaluation Metrics: For Event

- **Ranking Academic Event Related Search Result**
  - uses citations of committee members of events to rank search results of various documents.
  - it inherits the limitations and problems of citation measures.
  - Furthermore, it does not consider committee members’ roles in events.

- **Mining the characteristics of program committee members to discover prestigious conferences from collected data sets**
  - A good-quality conference has a comparatively higher number of PC members, papers, and co-authorships than one with low quality.
  - It does not determine which variables are the most contributing characteristics to determine the event’s quality.
2.2.1. Scholar’s Prominence

Core Group Type
- PC Member
- PC Chair
- OC Member
- Session Chair
- Invited Speaker
- Award Winner

Subject Field $F$

$W_{t1}$
- Core Group Type $t_1$

$W_{t2}$
- Core Group Type $t_2$

$W_{t3}$
- Core Group Type $t_3$

$W_{tn}$
- Core Group Type $t_n$

Scholarly Event 1
- Scholarly Event 2
- Scholarly Event 3
- Scholarly Event N

Toward Scholarly Event Digital Library Services
2.2.2. Scholar’s Prominence

\[
P(S) = \tau \frac{\sum_{t=1}^{n} (W_t K_t | F)}{(N | F)}
\]

the weight assigned to core group type \( t \)

the number of memberships for core group types in scholarly events

the total number of scholarly events in a given subject field \( F \)

Normalization factor to make \( P(S) \) constant

schorlar \( S_1 \) is said to be more prominent than scholar \( S_2 \)

if \( P(S_1) > P(S_2) \). \( Where, \ 0 \leq P(S) \leq 1 \)
2.2.3. Scholarly Event’s Prominence

Event $E_1$ is said to be more prominent than Event $E_2$ if $P(E_1) > P(E_2)$. Where, $0 \leq P(E) \leq 1$.
2.2.4. Scholarly Event Series’ Prominence

The total number of scholarly event instances belonging to Scholarly Event Series $\varepsilon$ in a given subject field $F$ is said to be more prominent than Event Series $\varepsilon_2$ if $P(\varepsilon_1) > P(\varepsilon_2)$. Where, $0 \leq P(\varepsilon) \leq 1$.

Subject Field Prominence of Scholarly Event $E$

Normalization factor to make $P(\varepsilon)$ constant

Event Series $\varepsilon_1$ is said to be more prominent than Event Series $\varepsilon_2$.

$$P(\varepsilon) = \tau \sum_{g=1}^{n} \frac{P(E)}{(G|F)}$$

the total number of scholarly event instances belonging to Scholarly Event Series in a given subject field $F$
2.2. What Stage is the project in?
Contents

1. Introduction
2. Current Work
3. Future Work Plan
   - Event Core Group Analysis
   - Scholarly Event Data Harvester
   - SNA based Information Platform
4. Research Implications

Toward Scholarly Event Digital Library Services
3.1. Event Core Group Analysis

• **Purpose**
  - To test the plausibility of the evaluation models in section 2.2,
  - To investigate whether core group membership in scholarly events can be used as a prominence index of scholars’ research achievement.

• **Hypothesis**
  - Core group membership has positive correlations with their better performance in citation measures (high correlation coefficient value) and
  - Scholars’ citation measures vary with Core Group Types.
### 3.1. Event Core Group Analysis

<table>
<thead>
<tr>
<th>Organizing Committee</th>
<th>Name</th>
<th>Affiliation</th>
<th>Country</th>
<th>Citation</th>
<th>h-index</th>
<th>Homepage</th>
<th>E-mail</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Chair</td>
<td>Jeffrey L.</td>
<td>University of Michigan</td>
<td>USA</td>
<td>2411</td>
<td>20</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td>President</td>
<td>Thomas J.</td>
<td>Auburn University</td>
<td>USA</td>
<td>195</td>
<td>5</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td>Tutorials</td>
<td>Jinyu Qi</td>
<td>University of California, Davis</td>
<td>USA</td>
<td>826</td>
<td>13</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td>Special Sessions</td>
<td>Robert J.</td>
<td>Carnegie Mellon University</td>
<td>USA</td>
<td>1200</td>
<td>14</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td>Exhibits</td>
<td>Paul J.</td>
<td>Imaging Research Laboratory</td>
<td>USA</td>
<td>286</td>
<td>13</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td>Finances</td>
<td>Jelena K.</td>
<td>Institute Informatics Lab, Carnegie</td>
<td>USA</td>
<td>869</td>
<td>50</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td>Publications</td>
<td>Miles W.</td>
<td>Department of Electrical and Computer</td>
<td>USA</td>
<td>869</td>
<td>50</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td>Local Arrangements</td>
<td>Murray H.</td>
<td>Electrical Engineering and Computer</td>
<td>USA</td>
<td>902</td>
<td>9</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td>Wiki Lewis</td>
<td>Venkat R.</td>
<td><a href="#">here</a></td>
<td>USA</td>
<td>500</td>
<td>8</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td>Review Committee</td>
<td>Till A.</td>
<td>Institute of Imaging &amp; Computer Vision</td>
<td>Germany</td>
<td>1030</td>
<td>12</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Purana A.</td>
<td>Medical Informatics, Medisys</td>
<td>Canada</td>
<td>1248</td>
<td>7</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Michael A.</td>
<td>Department of Ophthalmology</td>
<td>USA</td>
<td>628</td>
<td>13</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Scott A.</td>
<td>Department of Biomedical Engineering</td>
<td>USA</td>
<td>678</td>
<td>13</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Frances A.</td>
<td>Biomedical Imaging Group, Ecole</td>
<td>France</td>
<td>11</td>
<td>2</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Elsa A.</td>
<td>Signal &amp; Image Processing, Telecom</td>
<td>France</td>
<td>86</td>
<td>3</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Pascal B.</td>
<td>Department of Computer Science</td>
<td>Australia</td>
<td>26</td>
<td>1</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Peter B.</td>
<td>National Institute of Mental Health</td>
<td>USA</td>
<td>606</td>
<td>25</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Alberto A.</td>
<td>Department of Electrical and Computer</td>
<td>USA</td>
<td>247</td>
<td>4</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Peter B.</td>
<td>National Institutes of Health</td>
<td>USA</td>
<td>865</td>
<td>23</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Isabelle B.</td>
<td>Erasmus University Medical Center</td>
<td>The Netherlands</td>
<td>318</td>
<td>7</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Johan B.</td>
<td>Scientific Computing and Imaging Institute</td>
<td>USA</td>
<td>390</td>
<td>9</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Dana H.</td>
<td>INRIA</td>
<td>France</td>
<td>652</td>
<td>12</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Irene B.</td>
<td>Department of Radiology, University of</td>
<td>USA</td>
<td>1654</td>
<td>12</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Paul G.</td>
<td>Department of Aerospace and Mechanical</td>
<td>USA</td>
<td>629</td>
<td>16</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Carolina G.</td>
<td>Institute of Imaging and Biological</td>
<td>France</td>
<td>414</td>
<td>9</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Jean-Louis</td>
<td>INRIA</td>
<td>France</td>
<td>414</td>
<td>9</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Tim C.</td>
<td>Imaging Science and Biomedical E</td>
<td>France</td>
<td>2457</td>
<td>13</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Herve D.</td>
<td>INRIA</td>
<td>France</td>
<td>1118</td>
<td>17</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Silvane G.</td>
<td>Department of Biophysical and Electro</td>
<td>Switzerland</td>
<td>176</td>
<td>4</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Chair</td>
<td>Thomas S.</td>
<td>USA</td>
<td>195</td>
<td>5</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>F. De G.</td>
<td>INRA</td>
<td>France</td>
<td>1650</td>
<td>10</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Thomas M.</td>
<td>RWTH Aachen University, Department of</td>
<td>Germany</td>
<td>1200</td>
<td>14</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>Joelle D.</td>
<td>Leiden University Medical Center</td>
<td>Netherlands</td>
<td>59</td>
<td>4</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
<tr>
<td></td>
<td>J. D.</td>
<td>School of Medicine, Stanford University</td>
<td>USA</td>
<td>650</td>
<td>11</td>
<td><a href="#">here</a></td>
<td><a href="#">here</a></td>
</tr>
</tbody>
</table>

**Notes:**
- [here](#) indicates the absence of a direct link or address.
3.1. Event Core Group Analysis

<table>
<thead>
<tr>
<th>Plenary Speakers</th>
<th>Name</th>
<th>Affiliation</th>
<th>Country</th>
<th>Citation</th>
<th>h-index</th>
<th>Homepage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Biomedical Imaging</td>
<td>Elias A. Zerhouni</td>
<td>USA</td>
<td></td>
<td>7709</td>
<td>23 <a href="http://www.nih.gov/about/director/index.htm">http://www.nih.gov/about/director/index.htm</a></td>
</tr>
<tr>
<td>3</td>
<td>Imaging the Mouse</td>
<td>R. Mark Henkelma</td>
<td>Canada</td>
<td></td>
<td>5532</td>
<td>29 <a href="http://medbio.utoronto.ca/faculty/r.mhenkel@phenogenomics.ca">http://medbio.utoronto.ca/faculty/r.mhenkel@phenogenomics.ca</a></td>
</tr>
<tr>
<td>4</td>
<td>Ultrasound-enhanced</td>
<td>Katherine W. Fersi</td>
<td>USA</td>
<td></td>
<td>6291</td>
<td>22 <a href="http://femalab.bme.ucdavis.edu/efem@ucdavis.edu">http://femalab.bme.ucdavis.edu/efem@ucdavis.edu</a></td>
</tr>
<tr>
<td>5</td>
<td>Medical Image</td>
<td>Ron Kikinis</td>
<td>USA</td>
<td></td>
<td>11382</td>
<td>51 <a href="http://www.spl.harvard.edu/pages/rkikinis@bwh.harvard.edu">http://www.spl.harvard.edu/pages/rkikinis@bwh.harvard.edu</a></td>
</tr>
<tr>
<td></td>
<td>Computing</td>
<td></td>
<td></td>
<td></td>
<td>6563</td>
<td>31.25</td>
</tr>
</tbody>
</table>
3.1. What Stage is the project in?
3.2. Event Data Harvester

- **Harvester = Focused Crawler + Extractor**
  - Automatically identify scholarly events’ web sites and fetch their URLs, and then extract and add all found and relevant information to the SEDE knowledge base.

- **Some specific details regarding general information are often posted to scholarly event data listing services like EventSeer.**

- ‘Call for Papers' announcements are circulated via email or other forums.

- **However, the bulk of detailed information for events is typically posted on the event websites themselves.**
  - organizing committees, lists of topics, program schedule:

- **Details which are largely left out of alerting services.**
3.2. Event Data Harvester: Crawler

Start URL: http://www.google.com

Search String: molecular biology

Date Range: 2007년 1월 1일 부터 2008년 1월 1일 까지

Max Crawl: First Crawl Count: 10
Second Crawl Count: 15

Matches

<table>
<thead>
<tr>
<th>No</th>
<th>Title_Name</th>
<th>URL</th>
</tr>
</thead>
</table>

Status:

Contact
3.2. Event Data Harvester: Crawler

<table>
<thead>
<tr>
<th>No</th>
<th>Title Name</th>
<th>URL</th>
<th>Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Call for Papers</td>
<td><a href="http://ocs.sfu.ca/aish/index.html">http://ocs.sfu.ca/aish/index.html</a></td>
<td><a href="http://www.ieee.org/web/conferences/callforpapers/">http://www.ieee.org/web/conferences/callforpapers/</a></td>
</tr>
</tbody>
</table>

3.2. Event Data Harvester: I.E.

- **Extract bibliographic information (McCallum)**
  - extract bibliographic information from scientific articles, in order to build a computer science research paper directory automatically by using machine learning techniques

- **IE for call for papers (Shneider)**
  - Conditional Random Fields for the task of extracting conference names, titles, dates, locations and submission deadlines from CFPs.
  - the results show that the performance was at best 72.7%.

- **Automatic techniques clearly have the advantage that they can be run on large text corpora with little human effort,**

- **Automatic information extraction using NLP techniques have not seen very good results.**
3.2. Event Data Harvester: I.E.

- the style and format of scholarly events on the web comes in different forms.
- There is no standard for the posting of this varied information.
- Websites will often post this information on different kinds of pages, and in different contexts (different arrangements on the page and different typographical distinctions).
- Such information as dates, topics, and lists of people, are not presented in a standardized format, nor marked up in any way.
- Rather, the information is often 'floating' with only typographical clues as to their significance.
- All pages are designed in different ways and, coming from different countries of origin, tend to list information and use vocabulary differently.
3.2. Event Data Harvester: I.E.

- Due to the heterogeneous nature of event web pages, this project's initial strategy is to make use of a more 'low-tech' method of data extraction – utilizes manually defined patterns of text content and HTML formatting based on general conventions for listing data in human-readable formats on the web.
- Later machine learning algorithms will be investigated.

- Still however, machine learning approaches do not guarantee perfect automatic data extraction.

- As such, the current strategy is to use a semi-automatic approach.
3.2. Event Data Harvester: I.E.
3.2. Event Data Harvester: I.E.
### 3.2. Event Data Harvester: I.E.

<table>
<thead>
<tr>
<th>Person</th>
<th>Title</th>
<th>Affiliation</th>
<th>Email</th>
<th>Coun...</th>
<th>Committ...</th>
<th>Role</th>
<th>EventName</th>
<th>Misc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jack Y. Yang</td>
<td>General</td>
<td>USA</td>
<td>USA</td>
<td>Program</td>
<td>Chair</td>
<td>The 7th Int'l...</td>
<td>Jack Y. Yang</td>
<td></td>
</tr>
<tr>
<td>Distinguished Prof...</td>
<td>Pr...</td>
<td>University of Georgia, Athens</td>
<td>Co-Chai...</td>
<td>USA</td>
<td>Vice-Chai...</td>
<td>The 7th Int'l...</td>
<td>Georgia;</td>
<td></td>
</tr>
<tr>
<td>Massachusetts G...</td>
<td>Yuehui Chen</td>
<td>Massachusetts...</td>
<td>Co-Chai...</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Massachusetts;</td>
<td></td>
</tr>
<tr>
<td>Networks</td>
<td>Director</td>
<td>Director</td>
<td>Vice-Chai...</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Mississippi;</td>
<td></td>
</tr>
<tr>
<td>Southern Mississi...</td>
<td>USA</td>
<td>USA</td>
<td>and Sci...</td>
<td>USA</td>
<td>and Sci...</td>
<td>The 7th Int'l...</td>
<td>Mississippi;</td>
<td></td>
</tr>
<tr>
<td>Pattern Recognition</td>
<td>Northeaster University</td>
<td>USA</td>
<td>and Sci...</td>
<td>USA</td>
<td>and Sci...</td>
<td>The 7th Int'l...</td>
<td>Northeast...</td>
<td></td>
</tr>
<tr>
<td>Mary Qu Yang</td>
<td>Scientific Review</td>
<td>USA</td>
<td>and Sci...</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Yang;</td>
<td></td>
</tr>
<tr>
<td>Human Services</td>
<td>USA</td>
<td>Houston-clear lake</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Technology;</td>
<td></td>
</tr>
<tr>
<td>Computational Me...</td>
<td>ACCM</td>
<td>USA</td>
<td>and St...</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Book Seri...</td>
<td></td>
</tr>
<tr>
<td>South Korea</td>
<td>Osman Abul TOBB Unive...</td>
<td>USA</td>
<td>and St...</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Technology;</td>
<td></td>
</tr>
<tr>
<td>San Antonio</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Bi;</td>
<td></td>
</tr>
<tr>
<td>Southern California</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>California;</td>
<td></td>
</tr>
<tr>
<td>Central Florida</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Florida;</td>
<td></td>
</tr>
<tr>
<td>Hong Kong</td>
<td>Mark Clement Brigham Y...</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Clement;</td>
<td></td>
</tr>
<tr>
<td>Craig W. Codringt...</td>
<td>Jason Corso University</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Craig W.;</td>
<td></td>
</tr>
<tr>
<td>Detroit Mercy</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Detroit;</td>
<td></td>
</tr>
<tr>
<td>Massachusetts Lo...</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Massachusetts;</td>
<td></td>
</tr>
<tr>
<td>Massachusetts Lo...</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Massachusetts;</td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Massachusetts;</td>
<td></td>
</tr>
<tr>
<td>Wen-Lian Hsu Aca...</td>
<td>Taiwan</td>
<td>Taiwan</td>
<td>and St...</td>
<td>USA</td>
<td>USA</td>
<td>The 7th Int'l...</td>
<td>Taiwan;</td>
<td></td>
</tr>
</tbody>
</table>
3.2. What Stage is the project in?
3.3. SNA based Information Platform

1. Event Topics Distance/Similarity Calculation System

2. Research Trend Analyzing System

3. KOS Learning System

4. Academic Entities’ Networking System
3.3. SNA based Information Platform

1. Event Topics Distance/Similarity Calculation System:
   - In SEDE, event topics and session (track) topics are encoded in SKOS
   - Automatically measures and determines subject distances and similarities among events & event series
3.3. SNA based Information Platform

2. Research Trend Analyzing System
   - Enables users to analyze research trends,
   - overview continuously changing research topics, 
   - check hot and obsolete topics, and
   - converse among different disciplines
3.3. SNA based Information Platform

3. KOS Learning System:
   - Aims to build alternative Knowledge Organization Systems (KOS) for user-arbitrarily-specified subject (i.e., cell biology, medicine, mathematics).
   - Explore links to existing external KOSs (i.e., LexGrid, MeSH) to construe Knowledge Grid.
3.3. SNA based Information Platform

4. Academic Entities’ Networking System:
   - Delineates academic entities’ social relationships.
   - Researchers(groups)’ co-participations in a given events (series) will be analyzed to connect and group each other.
3.3. What Stage is the project in?

Diagram of project stages:

1. Preliminary Investigation
2. Detailed Investigation
3. Development
4. Validation
5. Commercial Launch

Key stages:
- Ideas
- Exploratory Research
- Development Research
- Technical Support

Flowchart showing the progression from ideas to commercial launch.
Contents

1. Introduction
2. Current Work
3. Future Work Plan
4. Research Implications

• Implications
4. Research Implications

- The proposed information platform can enhance library services by providing users with more comprehensive and organized scholarly information.

- SEDE will provide a background of EVENTOGRAPHY – An alternative field of bibliography or scientography.

- The Evaluation metrics will provide a background for the EVENTOMETRICS – A sub field of scientometrics.

- This research embodies the author’s hope to contribute to efficient scholarly communication.
Thank you!

Senator Jeong
senator@snu.ac.kr
http://www.eventography.org

http://bike.snu.ac.kr