

# Vision of Semantic Processing and the Latest Trends



---

**Dr. Nerute Kligiene,**  
[nerute@kti.mii.lt](mailto:nerute@kti.mii.lt)

Institute of Mathematics and Informatics,  
Vilnius Gediminas Technical University,  
Lithuania



# Introduction

---

- For centuries librarians have sought universal headings for storing and retrieving everything that has been written
- **Today the task is enlarged** to the level of
  - classifying, labeling, defining
  - finding, integrating
  - using everything on the World Wide Web



# Introduction

---

- The latest trends:
  - From finding relevant documents or content
  - A shift to **relationships** validating contextually relevant, meaningful relationships amongst the entities that documents mention and describe.
  - Semantic **metadata extraction** for automatic annotation, query processing and making inference

This is the next area of focus for the **Semantic Web community**



# Towards Semantic Web

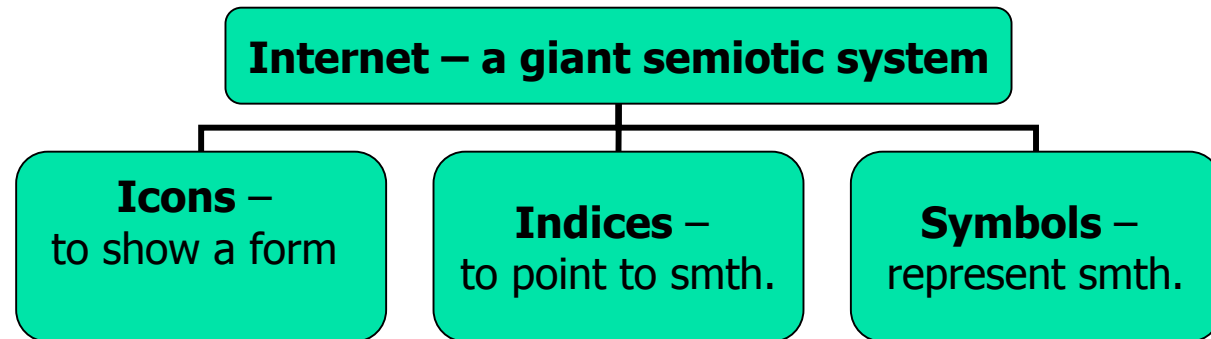
---

- The **World Wide Web** is rapidly becoming the universal repository for all the accumulated knowledge, information, data, and garbage of humankind.
- WWW >>> Metadata – a term covering many types of “**structured data about data**”



# Towards Semantic Web

---



Search engines harvesting for keywords, textual information but they were not able **to detect relationships** among them, **distinguish what was essential** and what was not.



# Some Basic Notions

---

- ***XML*** - ***Extensible Markup Language***, a flexible text format developed for WWW
- ***URI*** - ***Uniform Resource Identifier***, the generic term for all types of names and addresses referring to objects on the WWW
- ***URL*** - ***Uniform Resource Locator***, ***PURL*** – ***Persistent URL***
- ***RDF*** - ***Resource Description Framework***.  
RDF is a general framework for describing a Web site's metadata



# Emerging Semantic Web

---

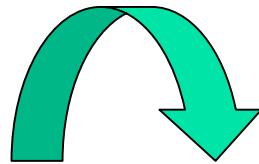
- The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is based on the **RDF** which integrates a variety of applications using **XML** for syntax and **URIs** for naming.
- "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." -- Tim Berners-Lee, James Hendler, Ora Lassila / Scientific American, May 2001



# New Trends

---

- **Online communities created by scientific publishers (*BioMedNet, Chem Web*): high-quality, current and authoritative material**



- **Traditional and long withstanding links between information professionals and users are being **altered!****





# New Means

---

- **Wide range of new materials **freely available** the Internet, without using the library:**
  - Pre-prints (e-prints)
  - Research manuscripts (working papers, technical reports, ...)
  - Electronic Thesis and Dissertations (ETD)



# E-Prints

---

- **E-print** - any electronic version of academic research manuscript circulated by the author outside of the traditional scientific publishing environment:
  - journal articles,
  - conference papers,
  - book chapters
- **“E-print archive”** is an online repository of these materials accessible to scholars



# E-Prints – Classification I.

---

- **Subject/discipline specific** and field wide *e-prints* repositories
- **Centralized** (subject based, based at single institutions)
- **Distributed** institutional e-print repositories



# **E-Prints – Classification II.**

---

- **Acceptance (or not) of non-referred papers**
- **Infrastructures to make research output available electronically**
- **Portals to facilitate access to domain specific grey literature**
- **E-prints repositories produced by commercial publishers**



# Benefits provided by E-Prints

---

## To Researchers:

- Increasing **visibility**
- Rapid **dissemination** of information to a wider audience
- Improved efficiency in the R&D activity and **faster communication** between academia and industry
- Improved **archiving** of scientific data – multimedia and supporting files
- E-prints offer **more features**, e.g annotation facilities/commentary by peers



# Benefits provided by E-Prints

---

## To Institutions:

- **Wider dissemination of R&D output**
- **Better reputation, possibility attract high quality researchers**
- **More chances to obtain research funds**



# Roles for Librarians in Self-Publishing by Researchers

---

- **Supporting** users to e-publish
- **Exploring** the document types accepted by e-print servers
- Providing **advice**
- **Increasing awareness** of possibilities and facilities provided by e-print archives
- **Persuading** institutional managers



# **Roles for Librarians in Self-Publishing by Researchers**

---

- **There is a great need to be involved into alliance among:**
  - **Librarians /information managers**
  - **Scholars, scientists and researchers**
  - **Those having responsibility for the development of infrastructures**





# Electronic Thesis and Dissertations (ETD)

---

- ***Networked Digital Library of Theses and Dissertations* – **NDLTD** since 1997 aims to:**
  - **Develop a federation of digital libraries,**
  - **Provide free access to graduate student's theses and dissertations;**
  - **Add to collaborative effort of universities around the world**
- **No cost for institutions interested to join *NDLTD*, <http://www.ndltd.org/join/>**
- **The *Guide for ETD* funded in part by *UNESCO* <http://etdguide.org/>**



# Electronic Thesis and Dissertations (ETD)

---

- **Roles for Librarians and Information Managers:**
  - **Advocate** to the Boards/Presidents/Rectors the potential of embracing *ETDs* and the advantages of being members of *NDLTD*
  - **Facilitate training** to faculty and students (as authors and supervisors)
  - **Study and propose solutions**, regarding archiving and preservation of the evolving *ETD* genre
  - **Complete the metadata** provided by *ETD* authors



# Conclusions

---

- **Information professionals have to be skillful in:**

- **Creating** and management of collections integrating resources in a variety of formats
- **Establishing links** between library catalogues and the new materials available on the Internet
- **Increasing scientist's awareness** of these new resources
- **Supporting** potential authors, by providing training on electronic publishing
- **Enhancing** user-created metadata



# Conclusions

---

- **(R)evolution in scholarly communication!**
- **Librarians have to meet a Wind of Changes**
- **Good Luck!**