INFS 427: AUTOMATED INFORMATION RETRIEVAL

Session 02 – Historical Developments in AIR

Lecturer: Mrs. Florence O. Entsua-Mensah, DIS
Contact Information: fentsua-mensah@ug.edu.gh

UNIVERSITY OF GHANA

College of Education
School of Continuing and Distance Education
2014/2015 – 2016/2017

Session Overview

The aim of this session is to:

- Provide an understanding of how the information retrieval field progressed/evolved to its current state.

  - Each evolutionary milestone is illustrated by describing the standards and protocols and by discussing the global initiatives and the research that shaped it.

Session Outline

The key topics to be covered in the session are:

- Topic 1: Information Retrieval Standards & Protocols
- Topic 2: Global Digital Library
- Topic 3: Intelligent Information Retrieval
- Topic 4: Hypertext and Hypermedia Systems

UNIVERSITY OF GHANA

Reading List

Rego, A., Garcia, L., Llopis, M., & Lloret, J. (2016). A New Z39.50 Protocol Client for Searching in Libraries and Research Collaboration. *Network Protocols and Algorithms*, *8*(3), 29. https://doi.org/10.5296/npa.v8i3.10147

The Apache Software Foundation: The Free and Open Productivity Suite. Retrieved from: http://www.openoffice.org/bibliographic/srw.html. Accessed on August 7, 2018.

UNIVERSITY OF GHANA

Introduction

- The growth of Information and Communication Technology (ICT) has refashioned information search and retrieval.

- There are several advancements that have taken place in this area over the period of time.

UNIVERSITY OF GHANA

Topic One

INFORMATION RETRIEVAL STANDARDS & PROTOCOLS

UNIVERSITY OF GHANA

What is a Standard?

- A standard means an agreement by what way to perform a task or carry out some activity to obtain a predictable result .

- There are various standards and protocols that are in existence for IR systems.

- Some of the popular search and retrieval standards and protocols include:
    - Z39.50
    - SRW
    - SRU
    - CQL

UNIVERSITY OF GHANA

Z39.50

- Z39.50 is a communication protocol between a client and a server.

- The increasing number of information available at libraries and the necessity to find a mechanism to look for information at several libraries at the same time promoted to the creation of the Z39.50 protocol.

UNIVERSITY OF GHANA

Z39.50 (Cont'd.)

- Sessions inside one connection between both nodes are known as Z39.50-association or Z-association (Rego et al., 2016).

- These sessions are initiated by the client. Since Z-association is open, both server and client can start any operation defined in Z39.50 protocol. In the same way, Z-association can be closed by either client or server, or implicitly terminated by loss of connection (Rego et al., 2016).

UNIVERSITY OF GHANA

Z39.50 (Cont'd.)

- The main goal of Z39.50 is to provide a standard to search information into an external database whatever its data organization.

- Thus, Z39.50 is widely used in some of the biggest libraries. This goal is achieved because the communication between the client and the server is standard and independent to the database (Rego et al., 2016).

UNIVERSITY OF GHANA

Z39.50 (Cont'd.)

- Z39.50 is used both at the national and international level as a standard protocol that defines computer-to-computer information retrieval technique. It is a non-proprietary and vendor-independent.

- Z39.50 was originally approved by the National Information Standards Organization (NISO) in 1988. In 1998, International Organization for Standardization (ISO) adopted Z39.50 and issued ISO 23950 Information and documentation - Information retrieval (Z39.50).

UNIVERSITY OF GHANA

Z39.50 (Cont'd.)

- Using Z39.50 a user through his/her system can search and retrieve information from other Z39.50 compliant computer systems without having the prior idea about the syntax of search that is used by the other systems.

- The primary goal of Z39.50 is to reduce the complexity and difficulties involved in searching and retrieving electronic information .

SRW

- SRW stands for **Search/Retrieve Web Service** protocol (The Apache Software Foundation, 2018). Its aim is to minimize the cross-language problems.

- The goal is to allow access to several networked resources and support interoperability among distributed databases, using a common utilization framework (The Apache Software Foundation, 2018).

- It is developed by collective implementers with more than 20 years of experience of the Z39.50 Information Retrieval protocol with nascent developments in the technological arena of the web.

UNIVERSITY OF GHANA

SRU

- SRU stands for **Search/Retrieve via URL**. It is a standard XML-based protocol for search by utilizing CQL (http://www.loc.gov/cql/), a standard syntax for query representation (The Apache Software Foundation, 2018).

- The prime difference between SRU and SRW is that the former uses HTTP as the transport mechanism and the latter is based on SOAP protocol and uses XML streams for both the query and the results.

- This depicts that the query is communicated as a URL and the XML is received as if it were a web page.

UNIVERSITY OF GHANA

CQL

- CQL stands for Contextual Query Language (formerly known as, Common Query Language).

- It is designed for use with SRW which is a search protocol successor to Z39.50 (as discussed in the previous section).

- CQL is an abstract and extensible query language for maximum interoperability amongst the connected systems. The goal is to reduce the difficulty to learn and use while retaining the capability to allow complex searches.

- Primarily CQL is used in the bibliographic domain, however it is not restricted to this context alone.

Topic Two

GLOBAL DIGITAL LIBRARY

UNIVERSITY OF GHANA

Global Digital Library

- This is more or less a virtual library that consolidates the collections of individual libraries as one collection.
  - The WWW and the internet laid the foundation for the virtual/digital libraries.

- Global Digital Library (GDL) is a prototype which aims to connect several national libraries and some major libraries, museums, archives, and information organizations with each other (Chen, 2001).

UNIVERSITY OF GHANA

Challenges/Issues with Digital information sharing

- Several legal issues may arise related to intellectual property, copyright, confidentiality and privacy, security, personal, business equity, etc.;

- Difference in culture may influence the way of information communication;

- The presence of generational gaps;

- The sheer complexity of information architecture both at the global and national level;

- To have an effective and adequate inventory of available resources comprising the knowledge of information;

- The ability to locate, identify and retrieve relevant and quality information;

- Due to the huge amount of information, the complexity arises related to "undesirable" "indecent" information.

UNIVERSITY OF GHANA

Topic Three

INTELLIGENT INFORMATION RETRIEVAL

UNIVERSITY OF GHANA

Intelligent IR defined

- Intelligent IR is a computer system having the capability to infer knowledge with the help of its previous knowledge for establishing a link between the requirement of its user and a set of candidate document (Jones et al., 2000).

- This is a system which can perform intelligent retrieval. The realization of researchers to use knowledge in the information retrieval system has led them to think about the artificial intelligent system which also has the similar purpose, and one among these classes is an expert system.

UNIVERSITY OF GHANA

Expert System Defined

- An expert system is "a computer system which emulates the decision-making ability of human experts" (Jackson, 1998).

- The expert systems are designed to solve complex problems by reasoning over knowledge stored in a knowledge base.

- The knowledge in the knowledge base is primarily represented as IF-THEN rules rather than conventional procedural code.

- The first expert systems were invented in the 1970s and then proliferated in the 1980s.

UNIVERSITY OF GHANA

Developments in Expert Systems

- As expert systems evolved, several new techniques were adopted into various types of inference, engines. Some of the most important ones include:
  - Truth Maintenance
  - Hypothetical Reasoning
  - Fuzzy Logic
  - Ontology Classification

UNIVERSITY OF GHANA

Expert Systems for LIS Profession

- **AUTOCAT** was produced in Germany. The system was designed to generate bibliographic records of physical sciences periodicals available in machine-readable form (Endres-Niggemeyer and Knorz, 1987) .

- **Qualcat** (Quality Control in Cataloguing) was undertaken at the University of Bradford. The goals of the project were to develop expert systems to select the best records, to link the databases and centralized authority control, to build a fully automated control package for day to day running, and to investigate interface problems for cataloguing (Ayres et al., 1994) .

UNIVERSITY OF GHANA

Expert Systems for LIS Profession

- **OCLC** developed an expert system, called Cataloguer's Assistant. The system was tested in Carnegie-Mellon University to reclassify the mathematics and computer science collection (De Silva, 1997) .

- **FRUMP**: (developed by DeJong) analyses articles from newspapers using frame-based techniques. The articles were first scanned and then data were automatically fed into the different slots within frames.

- **SCISOR**: (developed by Rau, Jacobs and Zernik, 1989) is a system that generate reports on corporate acquisitions and mergers.

UNIVERSITY OF GHANA

Topic Four

HYPERTEXT & HYPERMEDIA SYSTEMS

Hypertext

- Hypertext refers to the use of hyperlinks (or simply "links") to present text and static graphics. Many websites are entirely or largely hypertexts (Farkas, 2004).

UNIVERSITY OF GHANA

Hypermedia

- Hypermedia refers to the presentation of video, animation, and audio, which are often referred to as "dynamic" or "time based" content or as "multimedia" (Farkas, 2004).

- Hypermedia, a logical extension of hypertext, is a non-linear medium of information space which includes plain text, audio, video, graphics and hyperlinks link.

UNIVERSITY OF GHANA

Hypertext and Hypermedia (Cont'd.)

- Forms of hypertext and hypermedia include CD-ROM and DVD encyclopaedias (such as Microsoft's Encarta), eBooks, and the online help systems we find in software products.

- It is common for people to use "hypertext" as a general term that includes hypermedia (Farkas, 2004). For example, when researchers talk about "hypertext theory," they refer to theoretical concepts that pertain to both static and multimedia content.

(Farkars, 2004)

Summary

- In this session we have discussed some of the IR techniques and technologies that evolved in the recent past.

- We have discussed some of the significant IR standards and protocols.

- We have also reported the state-of-the-art research in IR field, for instance, the initiative of global digital library, application of intelligent systems like expert system in library cataloguing, classification and abstracting, the application and issues of intelligent hypertext and hypermedia systems.

UNIVERSITY OF GHANA

Activity 2.1

- Discuss the role of protocols and standards in the development of modern IR systems.

References - 1

Ayres, F. H., Cullen, J., Gierl, C., Huggill, J. A. W., Ridley, M. J., & Torsun, I. S. (1994). QUALCAT: automation of quality control in cataloguing. *BLRD REPORTS, 6068*.

Chen, C.-C. (2001). *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. Tsinghua University Press.

De Silva, S. M. (1997). A review of expert systems in library and information science. *Malaysian Journal of Library & Information Science*, *2*(2), 57–92.

Endres-Niggemeyer, B., & Knorz, G. (1987). AUTOCAT: knowledge-based descriptive cataloguing of articles published in scien-tific journals. In *Second International GI Congress 1987. Knowledge Based Sys-tems* (pp. 20–21).

UNIVERSITY OF GHANA

References - 2

Farkas, D. K. (2004). Hypertext and hypermedia. In *Berkshire Encyclopedia of Human-Computer Interaction* (Vol. 16, pp. 332–336). https://doi.org/10.1016/0360-1315(91)90062-V

Jackson, P. (1998). *Introduction to expert systems*. Addison-Wesley Longman Publishing Co., Inc.

Jones, K. S., Walker, S., & Robertson, S. E. (2000). A probabilistic model of information retrieval: development and comparative experiments: Part 2. *Information Processing & Management*, *36*(6), 809–840.

Rego, A., Garcia, L., Llopis, M., & Lloret, J. (2016). A New Z39.50 Protocol Client for Searching in Libraries and Research Collaboration. *Network Protocols and Algorithms*, *8*(3), 29. https://doi.org/10.5296/npa.v8i3.10147

The Apache Software Foundation: The Free and Open Productivity Suite. Retrieved from: http://www.openoffice.org/bibliographic/srw.html. Accessed on August 7, 2018.

UNIVERSITY OF GHANA