

# The Science and Art of Literature Searching for Computer Science Research

L. Gwenn Volkert, Ph.D.  
Department of Computer Science  
Kent State University

1

## Why do a Literature Search?

- To ensure you have a researchable topic.
- To ensure that you have enough material for writing a literature review.
- To provide a context for the research to be completed.
- To enable you to show evidence of both a broad and deep understanding your research topic.

2

## What You Need to Know

- How information is generated in your subject area
- How access points are created for this information
- How to organize your information
- How to analyze your information

3

## What are Information Sources?

3. **Tertiary** information sources
- ↳ 2. **Secondary** information sources
- ↳ 1. **Primary** information sources (i.e. the sources that you will cite in your report)

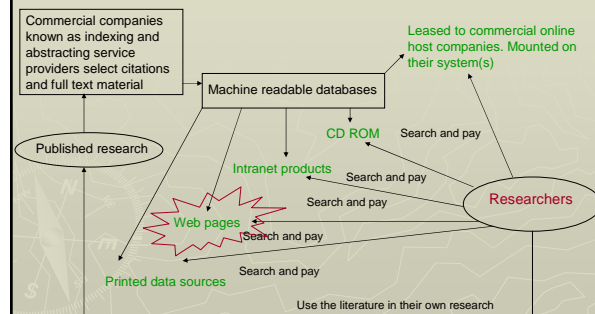
4

## Information Source Examples

1. **Tertiary**- course reading lists, indexing services, commercial databases, suggestions from your advisor
- ↳ 2. **Secondary**- text books, review articles, encyclopaedias, bibliographies
- ↳ 1. **Primary** articles in journals, conference papers, books, book chapters, dissertations

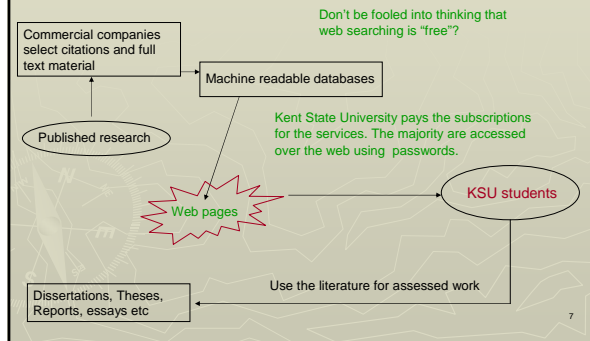
5

## Information Source Life Cycle



6

## Information Searching at KSU



## Indexing and Abstracting

- Locating sources of information on a topic of interest is difficult because of the magnitude of source documents
- Abstracts and indexes organize the literature into searchable databases so that a specialist can identify documents of interest more easily
- There are many different indexing and abstracting service providers to choose from

A nice introduction to I&A that is available at:  
<http://www.cas.usf.edu/lis/lis6260/lectures/indexabs.htm>

8

## I&A Service Providers



## Technical Aspects

- Indexing is the process of creating a searchable database
- No single database (I&A service provider) will cover all sources
- Learning to use keywords takes time and practice and becomes easier as you develop a better understanding of your topic.
- Many databases are available to KSU students through OhioLink

<http://www.ohiolink.edu/resources/dblist.php?by=subject&search=comp>

10

## Abstracting/Indexing Services

- ACM Computing Reviews, <http://www.reviews.com/>
- Science Citation Index®, (ISI)  
<http://www.isinet.com/products/citation/sci/>
- CompuMath Citation Index, (ISI)  
<http://www.isinet.com/products/citation/specialty/cmci/>
- Computer Source – EBSCO database, <http://search.epnet.com/>
- INSPEC – formally Science Abstracts, IEE,  
<http://www.iee.org/Publish/Inspec/>

11

## Abstracting/Indexing Services

- Computer Abstracts International-  
[http://beta.csa.com/factsheets/computab\\_s&t\\_cphp](http://beta.csa.com/factsheets/computab_s&t_cphp)
- Cambridge Scientific Abstracts:  
<http://www.csa.com/csa/index.html>
- Scopus – <http://www.info.scopus.com/>
- Mathematical Reviews- <http://www.ams.org/mathscinet>

12

## Other Abstract/Index Servers

- DBLP - Digital Bibliography & Library Project, Universitat Trier, <http://dblp.uni-trier.de> & <http://www.informatik.uni-trier.de/~ley/db/subjects.html>
- Citeseer - <http://citeseer.ist.psu.edu/>

13

## Why do I need to know this? I can just use search engines on the web!

- No single search engine indexes the data from all the primary sources that you might need
- Most web sites do not have keyword or subject descriptors
- Much of Internet material is "invisible"
- Where is the quality control?

14

## Journal Index Examples

The JOURNAL: *Parallel Processing*, Elsevier, is indexed by:

- ACM Computing Reviews
- Cambridge Scientific Abstracts
- Computer Abstracts
- CompuMath (ISI) -
- Current Contents/Engineering, Computing and Technology (ISI)
- Engineering Index
- INSPEC Information Services
- Mathematical Reviews
- Research Alert
- SCISEARCH (ISI)
- Science Citation Index (ISI)
- Scopus
- Zentralblatt MATH

15

## Journal Index Examples

The JOURNAL: *Journal of Algorithms*, Elsevier, is indexed by:

- ACM Guide to Computing Literature (ACM Portal)
- CompuMath Citation Index (ISI)
- Computer Contents (ceased)
- Mathematical Reviews (AMA)
- Research Alert
- SCISEARCH
- Science Abstracts
- Scopus

16

## Journal Index Examples

The JOURNAL: *Cluster Computing*, Springer, is indexed by:

- Computer and Information Systems Abstracts, (CSA)
- Electronics and Communications Abstracts,
- INSPEC Information Services
- Zentralblatt Math

17

## Journal Index Examples

The Journal: *Parallel Processing Letters*, World Scientific, is indexed by:

- INSPEC Information Services
- DBLP Bibliography Server
- Mathematical Reviews (AMA)

The Journal: *Soft Computing*, Springer, is indexed by:

- Current Contents (ISI),
- INSPEC Information Services

18

## Keywords

There is **NO** definitive list of keywords

The better you know your subject area of the literature, the better you'll know the keywords.

19

## Keyword Example

Search using **ISI Web of Science**  
(indexing 20,105,179 documents)

Keyword	Number of Documents Found
"Evolutionary Computation"	519
"Evolutionary Computing"	127
"Evolutionary Algorithms"	1057
"Genetic Algorithms"	6722

20

## More Keyword Examples

**ISI Web of Science:** (20,105,179 documents)

"Evolutionary Computation AND Hidden Markov Model"

- **No articles found**

"Evolutionary Computing AND Hidden Markov Model"

- **No articles found**

"Genetic Algorithms AND Hidden Markov Models"

- **6** documents matched the query.

"Evolutionary Algorithms AND Hidden Markov Models"

- **2** documents matched the query.

21

## And Another Keyword Example....

**ISI Web of Science:** (20,105,179 documents)

"Hidden Markov Models"-

**1142** documents matched the query.

"Hidden Markov Models AND Training"-

**207** documents matched the query.

22

## What About Other Databases?

Keyword search:

**"Evolutionary Algorithms AND Hidden Markov Models"**

- **PapersFirst** – 2 records found
- **ProceedingsFirst** – No records found!!
- **ArticalsFirst** – 1 record found
- **INSPEC** - 5 records found
- **COMPUTER SOURCE** - 2 records found

23

## A closer look at related terms

Two similar searches:

"Evolutionary Algorithms AND Hidden Markov Models"

**RESULT: 8 total papers**

"Genetic Algorithms AND Hidden Markov Models"

**RESULT: 6 total papers**

**But only 1 paper was duplicated in both results**

24

## Assessing Relevance

- This is a difficult topic and can easily lead to disagreements, even among established academics.

### Journals –

- Can you identify the leading journals covering your topic?
- Are the papers of sufficiently high quality?
- Are the members of the editorial board active researchers themselves?
- Does the journal have a reasonable impact factor?
- Is it a refereed journal

### Conferences –

- Can you identify the leading conferences covering your topic?
- What are the acceptance rates?
- What is the quality of the papers accepted?

25

## Assessing Author Relevance

### Authors –

- Identify the leading individual researchers/authors working in your area?
- What journals do they regularly publish in?
- Who is most cited (author impact)?
- Who publishes the most?
- Who produces students?
- Where do their students get jobs?

26

## More on Assessing Author Relevance

### Authors –

- Who gives the keynote talks and plenary talks at the major conferences?
- What, if any, editorial boards are they on? (the board members are listed in the journals)
- What conference program committees (PCs) are they on? (PC committees are usually listed on the conference website)
- What journals do they regularly review for (names are usually listed in the year end issue)
- What conferences do they regularly publish in?

27

## The Difference Between References and Citations

"The number of references a paper has is measured by the number of items in its bibliography as endnotes, footnotes, etc., while the number of citations a paper has is found by looking it up [in a] citation index and seeing how many others papers mention it."

quoted from **Price D. J. D.** *Little science, big science...and beyond*. New York: Columbia University Press, 1986. 301 p.

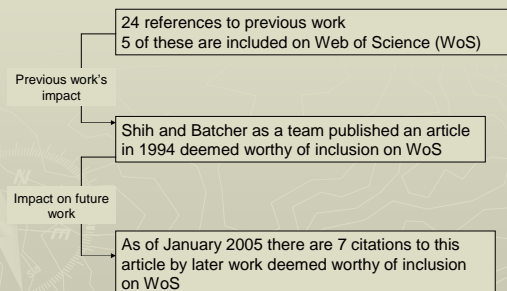
28

## Impact Analysis – cited reference searching (CRS)

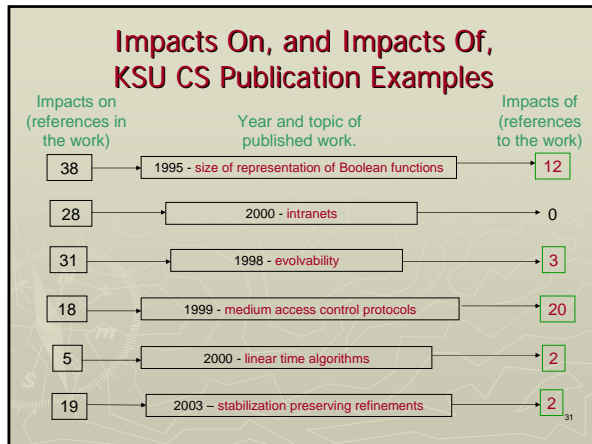
- Cited reference searching enables you to locate information on a given topic by using a representative or seminal work on the topic to search for other works on the same topic.
- The assumption is that articles that cite a seminal work must be related to it in subject matter.
- You can use CRS to follow the direction of research based on an earlier study.
- CRS can also be used as measure of the influence of a colleague's or competitor's work.

29

## Impact Example



30



## Grouping, Sequencing, Organizing and Documenting your Information

Personal bibliography database tools such as Bibtex, Endnote, ProCite are readily available

**Commercial:**

- Endnote website: <http://www.endnote.com/>
- ProCite: <http://www.procite.com/>
- Reference manager: <http://www.refman.com/>
- Biblioscape: <http://www.biblioscape.com/>

## Grouping, Sequencing, Organizing and Documenting your Information

Personal bibliography database tool such as Bibtex, Endnote, ProCite are readily available

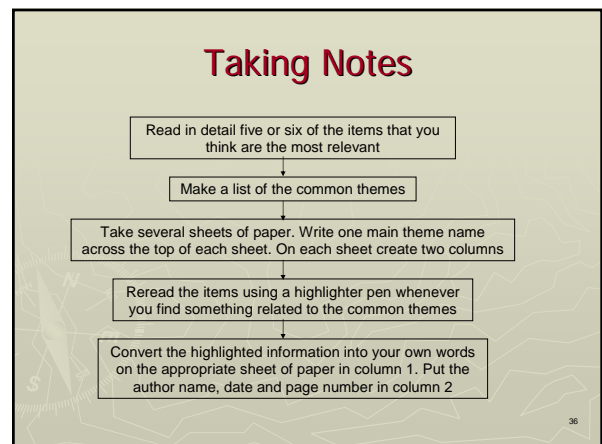
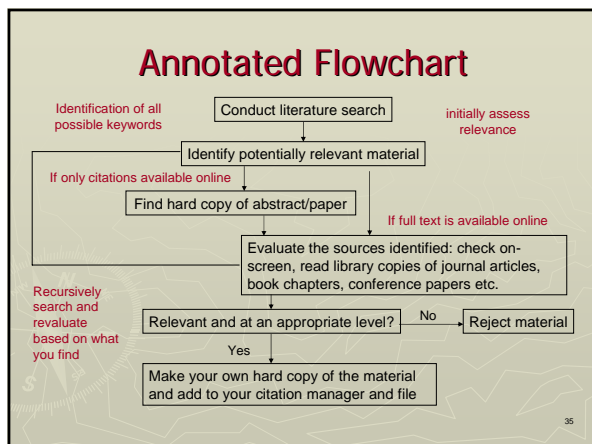
**Free:**

- RefWorks: <http://www.library.kent.edu/page/11033>  
(KSU provides an 8-campus subscription)
- Bibtex: <http://www.ecst.csuchico.edu/~jacobsd/bib/tools/bibtex.html>
- SixPack: <http://www.santafe.edu/~dirk/sixpack/>

## Organization Methods

Use a system for organizing material according to its relative importance, for example:

1. Number all print-outs of papers sequentially.
2. Store your papers in number order in a box or file cabinet.
3. Write the full citation across the top of each paper.
4. Write the paper numbers in an additional field for each paper's citation-manager entry



## An Example

Main theme

Sub-theme

7 main points on this topic on this sheet

This column shows the source of the main points

37

## Advantages of this System

- You are forced to think about what you are reading
  - When you record in your own words the material you have derived from others you are unlikely to accidentally quote another author - the work will clearly represent your interpretation of key issues
  - You have a record of where your information has come from, who said it, when that person said it and the page number for the information - helps with in-text citation work, checking back
  - You end up with a set of thematic notes
  - You can scale the system up for use with more material once you are confident that you have the main themes covered
- 38

## The Results of Your Efforts Are

- Clear evidence of a thorough literature search
  - Evidence of wide reading, high proportion of high level material
  - Evidence of understanding and assimilating what you have read
  - Thematic summary of up to date material
  - Discussion of trends in the discussion of the topic
  - Comments on the value of what has been studied
  - Citations will be formatted correctly in the text and in the listings
- 39

## Acknowledgments

- Dr. Hazel Hall, School of Computing, Napier University
  - ISI Web of Science
  - University of South Florida, School of Library and Information Science, College of Arts and Sciences
- 40