

Bioinformatics Colloquium Series
Spring 2005
John Grefenstette

- Schedule
- Requirements
- Writing Reports
 - Format
 - Ethical Issues
 - Style Issues
- Research Tools

Feb 1, 2005

1

Bioinformatics Related Program Faculty

Genomics

Patrick Gillevet
Curt Jamison
Don Seto
Jennifer Weller

Microbial Biocomplexity
Biomedical Genomics
Viral Genomics, Microarray Surveillance
Gene Expression Analysis

Protein Structure

Dmitri Klimov
Iosif Vaisman

Molecular Dynamics, Protein Folding
Protein Structure Analysis

Systems Biology

Saleet Jafri
Harold Morowitz
James Willett

Cellular Systems Modeling
Complexity, Origins of Life
Metabolic Models, Biodefense

Computational Methods and Data Mining

Dan Carr
Ken De Jong
John Grefenstette
Jason Kinser
Jeff Solka

Statistical Visualization
Machine Learning, Genetic Algorithms
Machine Learning, Network Models
Image Processing, Data Mining
Computational Statistics

Feb 1, 2005

2

Schedule (tentative)

- Feb 1 Administrative meeting
- Feb 8 John Grefenstette
- Feb 15 Jeff Solka
- Feb 22 Saleet Jafri
- Mar 1 Iosif Vaisman
- Mar 8 Dmitri Klimov
- Mar 15 No class (Spring Break)
- Mar 22 Jim Willett
- Mar 29 Pat Gillevet
- Apr 5 Harold Morowitz
- Apr 12 Curt van Tassell, USDA
- Apr 19 Ben Matthews, USDA
- Apr 26 Graduate Student Research Presentations

Feb 1, 2005

3

Requirements

- Two 4-5 page reports
 - On topic of one of the talks
 - Due 3/14/05 and 4/26/05
- Content
 - Short literature review on research topic
 - cite appropriate references
 - Summary of presentation
 - Summary of key open research areas

Feb 1, 2005

4

Writing Reports

- Use standard format
 - Title
 - Your name
 - Abstract
 - Background
 - Problem statement
 - Literature review
 - Methods
 - Results
 - Discussion
 - including open issues
 - References

Feb 1, 2005

5

Literature Review

<http://www.clet.ait.ac.th/EL21LIT.HTM>

- It is not supposed to be just a summary of other people's work!
- A critical look at the existing research that is significant to the work that you are carrying out.
- Although you need to summarize relevant research, it is also vital that you evaluate this work, show the relationships between different work, and show how it relates to your work.
- You need to select what parts of the research to discuss (e.g. the methodology), show how it relates to the other work
 - What other methodologies have been used?
 - How are they similar?
 - How are they different?
 - what is its relationship to your methodology?
- See sample in handout.

Feb 1, 2005

6

LITERATURE REVIEW: AN EXAMPLE

From <http://www.clet.ait.ac.th/EL21LIT.HTM>

On the optimal container size in automated warehouses

Y. Roll, M.J. Rosenblatt and D. Kadosh, Proceedings of the Ninth ICPR

Automated storage and retrieval systems (AS/RS) are being introduced into the industry and warehousing at an increasing rate. Forecasts indicate that this trend will continue for the foreseeable future (see [1]). Research in the area of AS/RS has followed several avenues. Early work by Hausman, Schwarz and Graves [6, 7] was concerned with storage assignment and interleaving policies, based on turnover rates of the various items. Elsayed [3] and Elsayed and Stern [4] compared algorithms for handling orders in AR/RS. Additional work by Karasawa et al. [9], Azadivar [2] and Parry et al. [11] deals with the design of an AS/RS and the determination of its throughput by simulation and optimization techniques.

Several researchers addressed the problem of the optimal handling unit (pallet or container) size, to be used in material handling and warehousing systems. Steudell [13], Tanchoco and Agee[14], Tanchoco et al. [15] and Grasso and Tanchoco [5] studied various aspects of this subject. The last two references incorporate the size of the pallet, or unit load, in evaluation of the optimal lot sizes for multi-inventory systems with limited storage space. In a report on a specific case, Normandin [10] has demonstrated that using the 'best-size' container can result in considerable savings. A simulation model combining container size and warehouse capacity considerations, in an AS/RS environment, was developed by Kadosh [8]. The general results, reflecting the stochastic nature of the flow of goods, are similar to those reported by Rosenblatt and Roll [12]. Nevertheless, container size was found to affect strongly overall warehousing costs.

In this paper, we present an analytical framework for approximating the optimal size of a warehouse container. The approximation is based on series of generalizations and specific assumptions. However, these are valid for a wide range of real life situations. The underlying assumptions of the model are presented in the following section.

LITERATURE REVIEW: AN EXAMPLE

From <http://www.clet.ait.ac.th/EL21LIT.HTM>

Notice how the writers have:

- grouped similar information: "Steudell [13], Tanchoco and Agee[14], Tanchoco et al. [15] and Grasso and Tanchoco [5] studied various aspects of this subject."
- shown the relationship between the work of different researchers, showing similarities/differences: "The general results, reflecting the stochastic nature of the flow of goods, are similar to those reported by Rosenblatt and Roll [12]."
- indicated the position of the work in the research area history: "Early work by Hausman, Schwarz and Graves [6, 7] . . . "
- moved from a general discussion of the research in AS/RS to the more specific area (optimal container size) that they themselves are researching i.e. they relate previous work to their own to define it, justify it and explain it.

Plagiarism and How to Avoid It

<http://www.indiana.edu/~wts/pamphlets/plagiarism.shtml>

- Plagiarism is using others' ideas and words without clearly acknowledging the source of that information.
- To avoid plagiarism, you must give credit whenever you use
 - another person's idea, opinion, or theory
 - any facts, statistics, graphs, drawings—any pieces of information—that are not common knowledge
 - quotations of another person's actual spoken or written words
 - paraphrase of another person's spoken or written words
- See URL above for examples of paraphrases that are actually plagiarism
- Plagiarism will be referred to the GMU Honor Council and cause failure for the course

Feb 1, 2005

9

Plagiarism and How to Avoid It

<http://www.indiana.edu/~wts/pamphlets/plagiarism.shtml>

Strategies for Avoiding Plagiarism

1. Put in quotations everything that comes directly from the text especially when taking notes.
2. Paraphrase, but be sure you are not just rearranging or replacing a few words.
3. Instead, read over what you want to paraphrase carefully; cover up the text with your hand, or close the text so you can't see any of it (and so aren't tempted to use the text as a "guide"). Write out the idea in your own words without peeking.
4. Check your paraphrase against the original text to be sure you have not accidentally used the same phrases or words, and that the information is accurate.

Feb 1, 2005

10

Citing References

http://www.indiana.edu/~wts/pamphlets/apa_style.shtml

All references appearing in Reference list shown be cited in the text of the paper

Typical text citations:

Several distinct classes of Boolean functions have been investigated, including random functions (Kauffman, 1969), canalizing functions (Kauffman, 1971), hierarchical canalizing functions (Szallasi and Liang, 1998), and Post classes (Shmulevich *et al.*, 2003).

Several distinct classes of Boolean functions have been investigated, including random functions [12], canalizing functions [13], hierarchical canalizing functions [15], and Post classes [22].

Typical Reference list item:

Kauffman, S. (1971). Gene regulation networks: a theory for their global structure and behaviors. **Curr Top Dev Biol** 6(6): 145-82.

[12] Kauffman, S. (1971). Gene regulation networks: a theory for their global structure and behaviors. **Curr Top Dev Biol** 6(6): 145-82.

Style in Science Writing

- Write in third person
 - Generally avoid "I", "we", "you".
- Cite results, not authors
 - Right: "DNA methylation is recognized as a main contributor to the stability of gene expression state (Feinberg *et al.*, 2001)."
 - Wrong: "Feinberg and his students show that DNA methylation is a main contributor to the stability of gene expression state (Feinberg *et al.*, 2001)."
- Tense
 - Past tense for experimental procedures:
 - "The solution was prepared using standard techniques."
 - "Algorithm A was run 20 times on each data set."
 - Present tense for algorithms or formulas:
 - "Algorithm A sorts the data and reports the smallest sequence."
- Learn by Reading
 - Read papers in journals, and notice the style

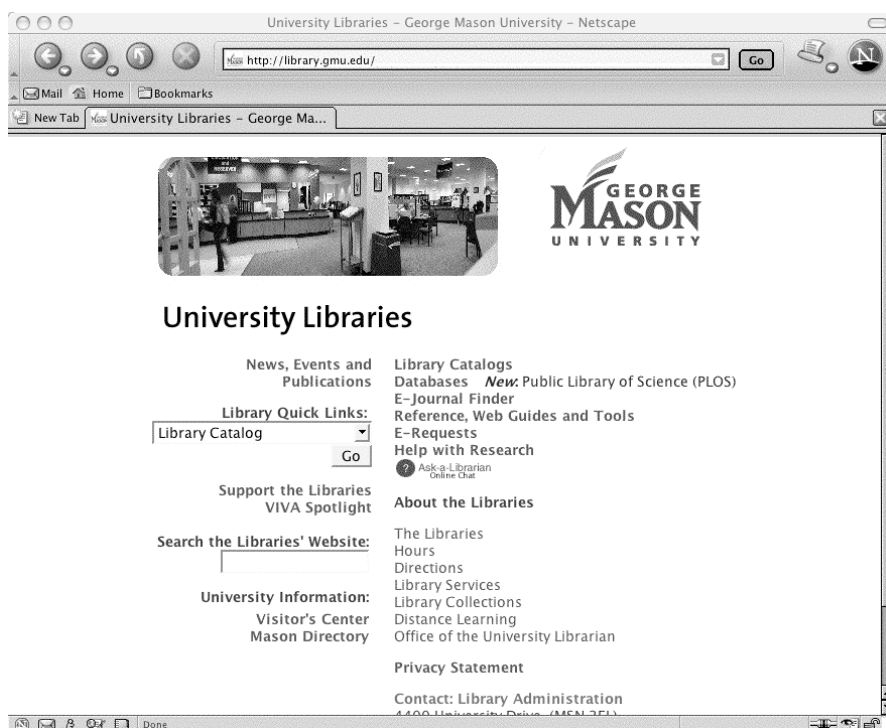
Research Tools

- Master a word processing program
 - MS Word, LaTeX
 - Learn to format figures, tables, and equations
 - Use spell-checker, grammar checker
- Search engines
 - Google
 - PubMed
- Learn to use the library resources at GMU
 - Electronic access to many journals
- Organize your literature reviews
 - EndNote, BibTex
- Master presentation software
 - Powerpoint, KeyNote

Feb 1, 2005

13

GMU Library

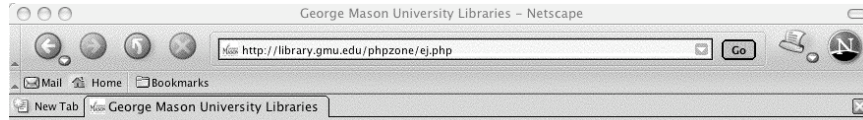


The screenshot shows a Netscape browser window displaying the GMU Library website. The browser's address bar shows the URL <http://library.gmu.edu/>. The website features a header with a photograph of a library interior and the George Mason University logo. Below the header, the main content area is titled "University Libraries" and includes several sections: "News, Events and Publications", "Library Quick Links" (with a dropdown menu set to "Library Catalog" and a "Go" button), "Support the Libraries VIVA Spotlight", "Search the Libraries' Website:" (with a search input field), "University Information" (listing "Visitor's Center" and "Mason Directory"), "Library Catalogs Databases" (listing "New: Public Library of Science (PLOS)", "E-Journal Finder", "Reference, Web Guides and Tools", and "E-Requests"), "Help with Research" (with an "Ask-a-Librarian Online Chat" link), "About the Libraries" (listing "The Libraries", "Hours", "Directions", "Library Services", "Library Collections", "Distance Learning", and "Office of the University Librarian"), "Privacy Statement", and "Contact: Library Administration (4400 University Drive, (MCN 2FL))". The browser's status bar at the bottom shows "Done".

Feb 1, 2005

14

GMU Library



e-journals

George Mason University

Mason users: How to access restricted databases from off-campus

Select by First Letter of Journal Title

0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Use the boxes below to check if a full-text, electronic copy of a journal is available. **When possible, the link will take you directly to the journal title.** In some cases you may have to navigate further once you reach the web site that is hosting the journal's content. If you wish to search for articles on a specific topic, please visit our database list.

SEARCH BY Journal Title (i.e. annals of botany)
SEARCH BY Journal Keyword (i.e. botany)
SEARCH BY ISSN (i.e. 0735-7044) SEARCH BY
PROVIDER (ie. JSTOR)

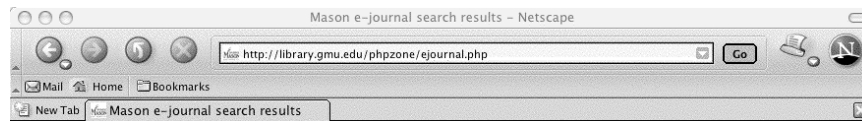
University Libraries Catalogs Databases Reference Research Web



Feb 1, 2005

15

GMU Library



e-journals

George Mason University

0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Often, the link below will take you directly to the journal of interest, but in some cases you may have to navigate further once you reach the web site that is hosting the journal's content.

Number of matches found: 5

Bioinformatics (Oxford, England)

1998 to present in Oxford Journals Online

Bioinformatics and Bioengineering, 2004. BIBE 2004. Proceedings. Fourth IEEE Symposium on

2004 to 2004 in IEEE Electronic Library Online

Bioinformatics and Biomedical Engineering (BIBE), Int'l Symposium on

2003 to 2003 in IEEE Electronic Library Online

Bioinformatics Conference (CSB), IEEE Computer Society

2002 to 2002 in IEEE Electronic Library Online

2003 to 2003 in IEEE Electronic Library Online



Feb 1, 2005

16

EndNote

- Free to GMU students
- Free GMU training seminars
- Organizes all your references
- Connects to PubMed and other online sites to search and download references
- Stores pdfs, abstracts, URLs
- Interacts with Word to insert citations in papers
- Easily format references in various styles for particular journals or conferences

