Background Research
Why Do Background Research?

• Learn an area — not everything is in your textbooks
  – Concepts, terminology
  – What are the important issues?
  – What are the main techniques?
    • e.g. modeling, simulation, proof techniques, statistics, algorithms

• Who are the important people in the area?
  – Where do they work?
  – What conferences do they attend?
Why Do Background Research?

• To find a problem to work on
  – What parts of a field are active
  – What do people consider to be today’s significant questions?

• To learn the customs of modern scientists
  – What kinds of papers: theoretical, analysis, simulation, implementation, experience reports.
  – What forms do they take?

• “Cultural” enrichment
Why Do Background Research?

• **To situate your own work**
  – What are the trends in the field?
  – How does your work differ from the work of others?
  – Are you going against the grain?

• **Avoid repeating the work of others**
  – Most forums for research require novelty
  – Except: survey papers,

• **Find out about the current “state of the art”**
  – Avoid getting scooped
Background Research

Finding relevant material
No single source that you can rely on
You have to use a combination of approaches
Not everything is on the web
not everything on the web is valuable – no quality control!
Direct Search

• Library
  – Books
  – Proceedings (newest material, shortest time delay)
  – Collections of papers (e.g., reprints, Comm. ACM’s first 25 years)
  – Journals
    • often have an annual index
    • on-line search of journal index very common
    • Some journals have back issues on line (ACM digital library)
  – Online access to electronic collections
Direct Search

• **Others’ offices, group libraries**
  – Who subscribes to what?
  – Who goes to which conferences?

• **Research group files**

• **I usually try to find a recent article first**
  Chaining backwards – can be easier than chaining forward
  CiteSeer: [http://citeseerx.ist.psu.edu/](http://citeseerx.ist.psu.edu/) is useful for chaining forward. (Also Science Citation Index — less good for conferences)
Indexing and Abstract Services

• **Current Contents**
  – Table of Contents are now often on-line

• **ACM Guide to Computing Literature**
  – Annual volumes

• **On-line indexing services**
  searchable via author, title, keywords
  sometimes include complete abstracts

• **Citation Indexes**
  – “Science Citation Index”
    Citeseer http://citeseer.ist.psu.edu/
  – limited to what they index
Organizing what you find

• At some point, I try to identify the earliest articles in a field
  – Sometimes these are the seminal works
  – Other times, later works are more organized and make a better introduction – e.g., survey articles

• What are the most often cited works?
  – Oft cited works are probably important works
  – Google.com ranks papers found in a search by how often they are cited
  – CiteSeer graphs citations by year

• Start trying to determine seminal papers, definitive references
  – How do the papers fit together?
  – Make a dependency graph of what papers cite others. This helps you view the development of a field
Leveraging What You’ve Already Found

- Bibliographies of books and articles already obtained
- Survey articles
  - ACM Computing Surveys
- Bulletins and notices
  - ACM SIGS
  - IEEE technical committees
  - bibliographies
  - research group archives
- IEEE Computer, IEEE Software
  - overview articles, for general audience
The Internet

• **Newsgroups**
  – DB/LP (database and logic programming)
  – comp.lang.functional (compilers/languages/functional)

• **Web pages**
  Research groups
  General sources — on-line bibliographies

• **Directories and search engines**
  National tech report library (NCSTRL)
  Computing Resource Repository (CoRR)
  [http://www.acm.org/corr/](http://www.acm.org/corr/)

• **Online Discussions (replace mailing lists)**
  – can provide pointers to more formal material
The Internet

• Direct solicitation
  On newsgroups
  In mailing lists and online discussions
  make sure that your questions are appropriate
  Email — give addresses for email and hard-copy mail
  Might ask if there is more recent work

If you have a technical report that is more than a few months old, it’s highly likely that it has been published someplace and probably revised
Miscellaneous

• Conference announcements
  – Advance copies of papers

• Manuals — some systems have no general publications

• Tutorial notes
  – usually from conferences, or summer schools

• Articles in encyclopedias, other references
  – tend to become outdated quickly
• **Notification services**
  – you subscribe to “keywords” and are notified when new works mentioning those keywords are published

• **Videotaped lectures**
  – Some libraries carry them
  – Increasing numbers now on the web

• **Laboratory annual reports**

• **Recent theses**
  – especially related work sections
  – UMI (formerly University Microfilms, expensive)
Overseas Research

Can be a real problem — often national journals and conferences tendency to miss this kind of work

Language barrier: most journals are not translated

important journals in German, French, Japanese, Russian, Chinese

Getting easier with the Web

many people post English descriptions of recent work

Google translate can give you an idea if work is relevant