Internet Investigations of Text Material to Compare Programs Across Institutions

Patricia B. Cerrito

Department of Mathematics

University of Louisville

Louisville, KY 40292

502-852-6010

pcerrito@louisville.edu

Introduction

We use web and text mining tools to examine institutional expectations concerning course syllabi and degree programs to demonstrate how such automated tools can be used to compare institutions. We focus our search on mathematics, and compare the results to other disciplines. The information we are seeking is readily available on university websites. However, it is difficult and time-consuming to extract as much of it as is contained within text documents. Also, each web site has its own method of locating the information through different link and retrieval options.

While only a small few have discussed the content of a mathematics PhD program and just a handful have discussed the content of a mathematics syllabus, there are few in the discipline who incorporate any suggested changes. We show that mathematics instruction remains very traditional, although some departments are moving to a more research oriented approach. This is quite contrary to the considerable changes that have occurred in K-12 mathematics education supported by the National Council of Teachers of Mathematics' standards of mathematics education.

In a paper published by the National Academies of Sciences (http://stills.nap.edu/html/grad/), there was a considerable discussion of graduate education, reflecting that the needs of industry should be considered. The PhD degree should encourage breadth as well as depth, and applied dissertations should be more readily accepted in PhD programs. A variety of sources, including the Boyer Commission on Undergraduate Education (http://naples.cc.sunysb.edu/pres/boyer.nsf/673918d46fbf653e852565ec0056ff3e /d955b61ffddd590a852565ec005717ae/$FILE/boyer.pdf) recommend that research should start at the undergraduate level and continue seamlessly through the dissertation. However, degree requirements do not appear to follow these recommendations.

Background

We want to look at differences in course syllabi in the sciences. We first examined syllabi at the University of Louisville and then extended the results to other science departments. Then we investigated across schools to examine similarities and differences in course syllabi. In particular, we wanted to determine whether there was a regular consideration of alternative learning styles, the use of writing, or the use of technology.

Very typical in a mathematics syllabus is the chapters in the book covered, homework, exams, and grading. Consider this information from a trigonometry syllabus (Display 1).

Display 1. Material Covered in Trigonometry

Material We will cover the first four chapters and several sections from Chapter 6, time permitting. This is an enormous amount of material to cover in 42 class periods. Because of this, not everything will be covered in class. Don't assume material not covered during lecture is not required! You have to read the book!

Mathematics is concerned with solving real-world problems through mathematical models. This course satisfies the general education requirement in the content area of mathematics.

Consider a second display, this time for a course in Finite Mathematics.

Display 2. Material Covered in Finite Mathematics

Text: FINITE MATHEMATICS by Barney, Ziegler, and Byleen; 11th Edition (special for UofL).

Material: This is a first course in finite mathematics, taken mainly by students majoring in business, but also by students wishing to fulfill the quantitative reasoning (math) general education requirement. The material consists of: Systems of linear equations, Gauss-Jordan elimination, matrix operations (addition, multiplication, inverses, matrix equations), linear inequalities, set operations (union, intersection, complement), probability (conditional probability, Bayes rule. Binomial and normal distribution), Markov chains (regular chains, absorbing states); basically Chapters4,part of 5, part of 7, 8, 9, and part of 11 in the current text.

Even though the two courses cover very different material, the approach is the same; what parts of the book will be covered. We next consider course grading (Displays 3-5).

Display 3. Course Grading Using Exams and Quizzes

Grading: Three tests (see schedule below), one after each chapter; at least twelve quizzes (unannounced, but mostly Fridays), only 10 best quizzes will count; comprehensive final exam. Grades for this course will be based upon total points earned from the quizzes, final exam and tests. Each of the three semester tests and the final exam is worth a maximum of 100 points, each quiz 10 points (for a total of 100 points. This gives a total of 500 points.

Out of 500 points,

480 is required for A+, 450 is required for A, 440 is required for A-,

430 is required for B+, 400 is required for B, 390 is required for B-,

380 is required for C+, 350 is required for C, 340 is required for C-,

330 is required for D+, 300 is required for D, and 290 is required for D-.

There will be NO make-up quizzes, if you missed a quiz with an approved excuse, you can drop it. No late work will be accepted. If you miss an exam with an approved excuse, arrangements for a make-up exam need to be made prior to missing the exam (only in the case of severe problems will a make-up be granted later).

Display 4. Course Grading Using Quizzes Only

Grading There will be no hour tests in this course and homework will not be collected. Instead, the grade will be determined from total points earned from weekly quizzes, extra credit and a final exam.

Quizzes The first 15–20 minutes of each class period on Wednesdays will be used for a 30 point quiz. The problems on each quiz will be very close copies of homework problems assigned earlier in the semester.

Display 5. Course Grading Using Homework and Participation

GENERAL INSTRUCTIONS TO STUDENTS. Attend all lectures (required). Read the material to be covered prior to coming to class. Make a list of questions on topics that need clarification and address these during the lecture. Work the Homework assignments listed on the attached schedule. You are required to do all the homework problems and homework will be collected and spot checked. Ask questions about the homework during class or see me during office hours. You may present homework problems in class at the board or as a Maple worksheet (1 point each, counts as participation credit). Homework problems may be collected as parts of quizzes, **so you must have homework from sections covered within the previous two weeks with you during lectures**. Unannounced quizzes or other exercises worked for a grade may be given during any class. Tests may be in class or takehome. The number and timing of tests, as listed in the tenative schedule posted on the course web site is subject to change as the semester progresses. Read the Course Announcements page regularly (at least twice per week) for current information about the course, assignments and due dates. Participation credit is available for contributions to the class, such as posing (or answering) questions, creating lecture outlines or MAPLE worksheets and submitting them in a form suitable for posting.

GRADES: Your grade will be based on the following weighted components:

- 80% Test (at least three), quizzes and special problem sets.
- 10% Homework is collected and spot checked.
- 10% Component based on participation points. Attendence is required and forms a part of the participation grade (see participation points below).

The plus/minus grading scale will be used in this class. The following percentage are required for each grade: 92=A, 89=A-, 86=B+, 83=B, 80=B-, 77=C+, 74=C, 71-C-, 68=D+, 65=D, 62=D-, 61=F .

Note that Displays 3 and 5 give points and/or percentages for grades with a weighting for each type of assessment; Display 4 does not give such specific information. The final component of the syllabus is a list of homework problems (Display 6).

Display 6. Homework List

Day	Section	Homework problems: odd in given range
1. 1/8 (T)	Introd. Plus 4.1	3,7,11,15,17,25,29,51-59,63
2. 1/10 (R)	4.2	7-19,43-65
3. 1/15 (T)	4.3	9,11,19,25,29,33,37,43,47,59,67-71
4. 1/17 (R)	4.4	1-17,23-43,51,61
5. 1/22 (T)	4.5	9-17, 23, 35, 39-43, 57, 59, 65, 67
6. 1/24 (R)	4.6	3, 5-21, 23-29
7. 1/29 (T)	4.7	1-5, 13, 17,19,23-31
8. 1/31 (R)	Review	

Occasionally, there is a list for student responsibilities (Display 7).

Display 7. Student Responsibilities in Class

Student Responsibilities:

1. **Read the book before class**
2. Write down questions
3. Attend ALL lectures
4. Be on time
5. Stay for entire class
6. Check with other students if you had to miss a class
7. Do all the assigned homework
8. Read, prepare, solve problems for about 6 to 9 hours every week
9. If you have special needs (disability, participation in athletic activity, etc.) let me know immediately, to make arrangements.

Syllabi in mathematics tend to be very brief, usually one page with a second page containing a list of homework problems.

Display 8 shows an excerpt from a biology course. It is briefer still, but also gives a list of homework topics covered rather than homework problems (Display 9).

Display 8. Biology Syllabus

• **Expectations and course mechanics:** This class is targeted towards biology seniors. Students are expected to use suitable resource materials, either the recommended book or other books written for (micro)biology majors to study the biology of bacteria. Class attendance is recommended but not required except for the scheduled exams. Professional attitude and behavior in class is expected. It is the students' choice to enroll in or dis-enroll from the class; however, a petition to the pertinent A&S committee might be required. Learning progress will be assessed by one midterm exam (multiple choice) and a comprehensive final exam (written essay or oral). **Passing of the comprehensive final exam is required for passing the course.** Exam questions will cover material presented in lectures and assigned reading. The midterm exam will be worth 40%, the final exam contributes 60%. Students can elect to do a term project worth 10 %, which will be figured into the overall score. Grading will be no stricter than: A: > 90%, B: 80 - 89.9%, C: 70 - 79.9%, D: 60 - 69.9%, F: < 60%. In order to take a make-up exam for full credit, arrangements must be made **before** the scheduled exam date (except for documented **emergencies** or **University-sponsored absences**). If an exam falls on a day on which the University is unexpectedly but officially closed, the exam will automatically be given during the next regularly scheduled class period. Students are responsible for obtaining information on class/campus closures if such an event occurs. Students will have the choice of taking a written or oral final comprehensive exam. This choice will be binding and documented in writing on or before April 20, 2008.

Display 9. Topics Covered

TENTATIVE SCHEDULE OF CLASSES (DA203; T R 11:00 – 12:15 PM):

	Date		Subject
January	8	T	Introduction/ The Prokaryotic Cell
	10	R	The Prokaryotic Cell
	15	T	The Prokaryotic Cell
	17	R	Nutrition & Microbial Growth
	21	*M*	*Dr. Martin Luther King, Jr. Day*
	22	T	Nutrient Assimilation
	24	R	Metabolic lifestyles (ML): The diversity of catabolism – an introduction
	25	*F*	*Last day to apply for degree*
	29	T	ML: Chemoorganotrophy
	31	R	ML: Chemolithotrophy
February	5	T	ML: Phototrophy
	7	R	ML: Autotrophy
	12	T	ML: The uniformity of Biosynthesis
	14	R	ML: Biosynthesis
	19	**T**	**Midterm Exam**
	21	R	Communication with the Environment (CE): Introduction (2 component systems)

In a more advanced course in biology, both oral and written communication are introduced (Display 10). As such, the syllabus contains information on the required writing (Display 11).

Display 10. Advanced Undergraduate or Graduate Course

Course Components: I will be using a variety of experiences to help you build your ecological understanding and to help me assess your grasp of the objectives described above. Below are the basic components. Everyone will be graded on a straight percentage basis: 90-100% is an A, 80-89% is a B, 70-79% is a C, 60-69% is a D, and < 60% is an F. Plus and minus grades will be given within these ranges.

25%	Exam 1
25%	Exam 2
20-35%	Final Oral Presentation
20%	Written Final Paper (Graduate Students)
15-20%	Paper Discussions
5%	Class Participation
100%	

Display 11. Writing Instructions

Final Oral Presentation: The Final Oral Presentations will be scheduled during the last weeks of class (see syllabus for dates). The topics for these presentations are to be selected and approved by me by **Tuesday, March 18.** You are free to choose as a subject for your presentation any topic that relates to the course content - Population and/or Community Ecology – but it should not simply be a summary of your current or past research. Ideas you are considering for future research are fine. I will arrange the schedule of the presentations so that talks on similar topics are presented on the same day. Your presentation will be for a total of **15 minutes – including questions!** Plan to have your talk be 10-12 minutes, so that there is time for questions.

On the day of your presentation you are expected to turn in to me and have copies for each person in class (total ~25):

    1) An Abstract or Summary of your Talk (1-2 paragraphs; 1/2 page maximum)
    2) A list of references you used for your talk (minimum of 5)

Your final oral presentation should be given in the format of a contributed paper at a scientific meeting. I will base the evaluation on how well you present and communicate the main points, use visual aids to support your ideas, and the quality and appropriateness of the references you use to develop the presentation. I suggest the following be used as guidelines in preparing your presentation.

    State the question – What is the topic of your talk and why is it interesting/important?

    Conceptual framework – What theory/concerns/observations are motivating this work?

    System – What will you be talking about? - What is the focal species, system, habitat, or theory… give some background on the system of interest.

Next, we turn to chemistry. These syllabi generally integrate lab reports and lab requirements with exams (Display 12).

Display 12. Grading for Chemistry

Grading: Your grade will be based upon your performance on weekly experiments and on the final exam. The experiments will contribute 3/4 of the grade, and the final exam will contribute 1/4. The lab report consists of **both** the pre-lab questions and the actual lab report, both of which can be found in the laboratory manual. The pre-lab question sheet must be completed by each student and approved by the TA at the beginning of the lab before you will be allowed to begin the scheduled experiment. Your laboratory TA will instruct you as to when the reports are due.

		Grading Scale:	
Lab Reports (100 points each)	– 600 points	800-720 pts (90%)	A
Lecture Exam –	200 points	719-640 pts (80%)	B
	800 total points	639-560 pts (70%)	C
		559-480 pts (60%)	D
		479 or less (<60%)	F

However, information on the lab reports is very sketchy (Display 13).

Display 13. Information in Syllabus on Lab Reports

A schedule of laboratory experiments is included in the syllabus. You must read each experimental procedure **BEFORE your lab session.** A set of pre-lab questions must be answered prior to beginning the experiment. Make sure you (1) know the purpose of the experiment, (2) understand the concepts and procedures to be conducted, and (3) are aware of any safety concerns and special waste disposal requirements. You must be familiar with and abide by all safety procedures posted in the lab room and detailed in the laboratory manual. Proper clothing and safety goggles are required at all time; **students will not be admitted to the lab if clothing and goggles requirements are not met!**

When a lab is not part of the course, the course grading depends entirely upon examination (Displays 14).

Display 14. Chemistry Course Without Lab

The course grade is based on three semester exams (1/6 each), a recitation grade (1/6), and a Final Exam (1/3); the Final Exam is comprehensive. Dates for all exams are announced in advance and must be taken during the scheduled times. THERE ARE NO MAKE-UP EXAMS. A missed semester exam will count as zero and will be calculated into the course grade as such; a missed Final Exam is an automatic F for the course. There are some cases where a student will miss an exam for justifiable cause, such as illness, surgery, etc. In these cases, arrangements can be made so the student is excused and is not penalized with a zero. These arrangements must be made with Dr. Noble and should be done in advance of the exam date if possible. If this is not possible, then it is the student's responsibility to contact Dr. Noble promptly thereafter. (A routine appointment for a routine reason, medical or otherwise, is not considered justifiable cause if it can be conveniently scheduled at another time.) Generally, requesting an excuse for justifiable cause will require some written evidence such as a doctor's back-to-work/school statement, etc. Some excuses are not considered justifiable; examples include traffic, dead car battery, defective snooze alarm, etc.

We turn to Physics (Display 15,16). The syllabi tend to be as brief as the other sciences.

Display 15. Physics Course Requirements

Requirements

This course is based on the astronomy textbook, *The Essential Cosmic Perspective*, Second edition by Jeffrey Bennett, Megan Donahue, Nicholas Schneider, and Mark Voit. A brief quiz will be given each friday on the sections of this book to be covered during the week and during the following week. The purpose of the quizzes is to make sure you keep up with the reading and are well-prepared for the discussion in class. On each friday, I will say which sections of the book will be covered on the following quiz. In addition to the required text, I recommend the book, *Night Watch, A Practical guide to Viewing the Universe* by Terence Dickerson, which is a helpful guide for learning the night sky. The course will include three multiple choice tests. These tests will be given on

Test 1 Friday, February 13
Test 2 Friday, March 26
Test 3 Monday, April 26

Please plan to attend each exam. Unless a student has a verified excuse, a missed exam will count as a zero. The cumulative score students receive on the quizzes will be given the same weight as the three tests in determining their grade. The course website is *http://www.physics.louisville.*

Display 16. Second Course in Physics

Textbook: Optics, Eugene Hecht, 4th ed., Addison-Wesley. Some sections of the textbook will be omitted due to time constraints.

Evaluation:
- 3 exams (2 mid-terms and 1 final-exam) equal weight for each: 75%
- Homework (approximately 10 assignments): 25%
- Quiz (extra-credit): 10%

Grading:
- A = 100% - 85%
- B = 84% - 75%
- C = 74% - 65%
- D = 64% - 55%
- F = below 54%

Displays 1-16 show the general trend in terms of syllabi in the sciences. They are brief, usually listing the textbook material to be covered. Most of the grade is based upon examination, although occasionally homework is graded. Rarely will participation in class or communication be included in the course. This, then, is the trend at the University of Louisville. We wanted to compare to other universities.

Method

The first step in the process of investigation is to perform a keyword search on the term, "mathematics syllabus and university". Because of the preponderance of advertisements that now appear at the beginning of a returned search, we need to include enough search terms to minimize the appearance of ads. Then, the SAS macro, %tmfilter, is used to collect the returned documents into a directory,

```
%tmfilter(url=http://www.ask.com/web?q=mathematics+syllabus+and+univ
ersity&sm=adv&qsrc=196,
   depth=1,
   dir=c:\syllabi\dir,
   destdir=c:\syllabi\destdir,
   norestrict=1,
   dataset=syllabi.documents);
```

Once the documents are placed in the directory (c:\syllabi), the document locations are stored in a SAS dataset, syllabi.documents. These locations are used to point to the directory so as to conserve memory storage. We did a

second search to investigate mathematics PhD degree requirements using the search URL,
http://www.ask.com/web?q=mathematics+phd+degree+requirements&search=search&qsrc=0&o=0&l=dir.

Once the web sites have been gathered into a document folder, they are analyzed using the process of text parsing and singular value decomposition. The purpose of these two steps is to remove terms from the documents that do not contribute to any meaningful analysis, terms such as "of", "and", "the". Parsing then involves creating a term by document matrix. As this matrix is extremely large, singular value decomposition compresses the matrix to a smaller size, say 100 dimensions. This compression can be done since the term by document is very sparse (containing mostly zeros).

Text mining uses both grammar and syntax to find documents that are similar. Different weights are given to different words (with the most commonly used terms given the least weights), and the weighted terms are used to define the documents into mutually exclusive clusters.

Results

Course Syllabi

For the course syllabi, almost 300 documents were returned. Documents from the ads at the beginning of the returned search were filtered out of the analysis. What documents remained were defined into a total of 6 groups. The term "exam" appeared 1282 times in 161 documents, and the similar term "quizzes" appeared 353 times in 102 documents. "Test" appeared 911 times in 146 documents. The term "grade" appeared 1154 times in 174 documents. In contrast, "homework" appeared 698 times in 136 documents, indicating that many of the syllabi did not discuss homework, although the term "assignment" appeared 503 times in 135 documents. Figure 1 gives the clusters of documents of the mathematics syllabi.

Figure 1 shows which groups of documents focus on which items. For example, cluster 5 with 69 documents is very focused on exams and homework, as is cluster 6 with 31 documents. In contrast, cluster 1 is more focused on course content, discussing concepts such as polynomials and distributions. Documents in cluster 3 spend time on accommodations for students with disabilities. It suggests that some of the syllabi in clusters other than 3 might need to improve their information on disability accommodations.

Figure 1. Clusters of Documents

| #	Descriptive Terms	Freq	Percentage
1	+ square, + distribution, + property, + variable, + equation, + design, + polynomial, linear, + apply, + computer, basic, + system, + concept, + model, + function, + credit, + number, + method, + graph, + description	31	0.113138686...
2	+ appear, + drop, + withdraw, + receive, + college, + find, + need, + question, + date, + write, + make, + have, + material, + test, + include, + do, + objective, + other, + assignment, + graph	34	0.124087591...
3	+ last, + contact, + revise, technological, + topic, + university, tennessee technological university, + disability require accommodation, + reference, + center, first, + request, + accommodation, + possible text, roaden university center, department, + end, + locate, catalog, + prerequisite	57	0.208029197...
4	+ graph, + provide, + write, + date, + learn, + solve, + policy, + syllabus, + function, + make, + grade, + topic, + other, + have, + number, + calculator, + assignment, + hour, + final, + time	52	0.189781021...
5	+ exam, + schedule, + grade, + final exam, + final, + miss, + homework, + not, + material, + class, + give, + will, + section, + assignment, + hour, all, + day, + syllabus, + do, on	69	0.251824817...
6	+ meet, + need, + review, + email, + question, + policy, + follow, + assignment, + day, + do, + time, + have, + provide, + work, + exam, + grade, + study, + syllabus, + hour, + complete	31	0.113138686...

Figures 2-6 show concept links, which demonstrate how terms in the syllabi are linked. Figure 2 is centered at the term, "final exam" as it is linked from the term, "exam". Of primary interest is that some of the syllabi indicate that hand-held calculators are acceptable for examinations.

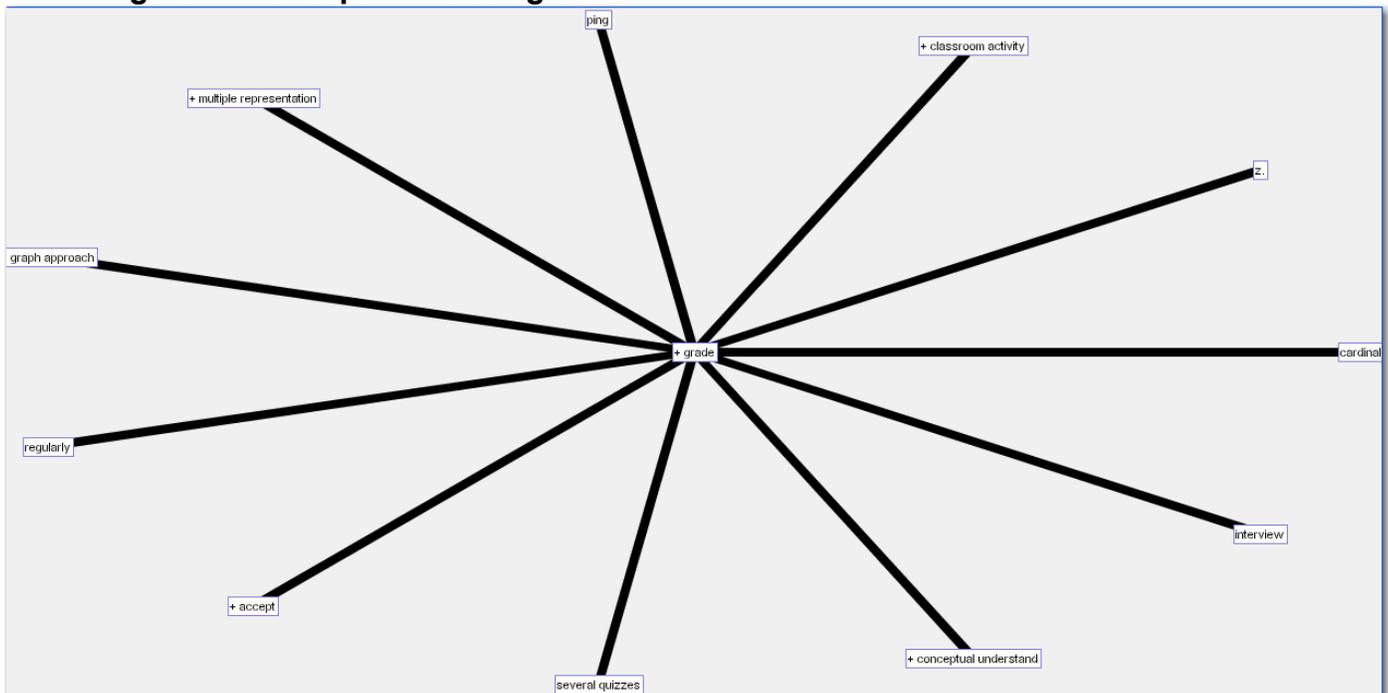**Figure 2. Concept Links to "final exam"**
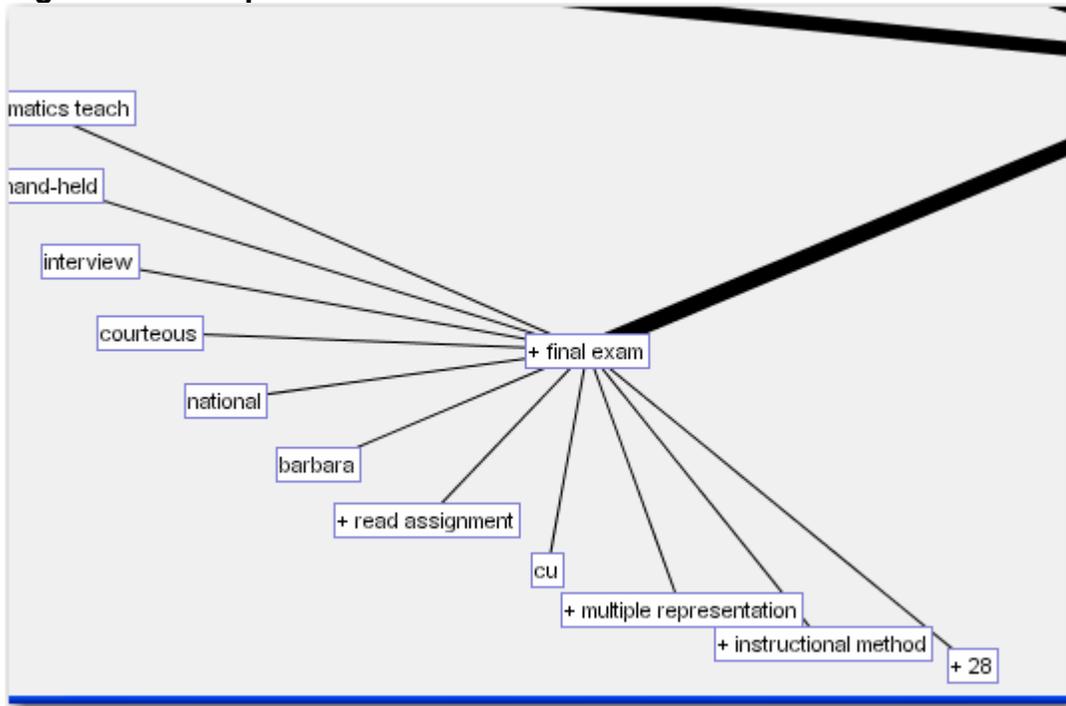
Two terms of interest connected to "quizzes" is "mastery" and "instructional".
Documents in these links emphasize regular quizzes as part of course grade.

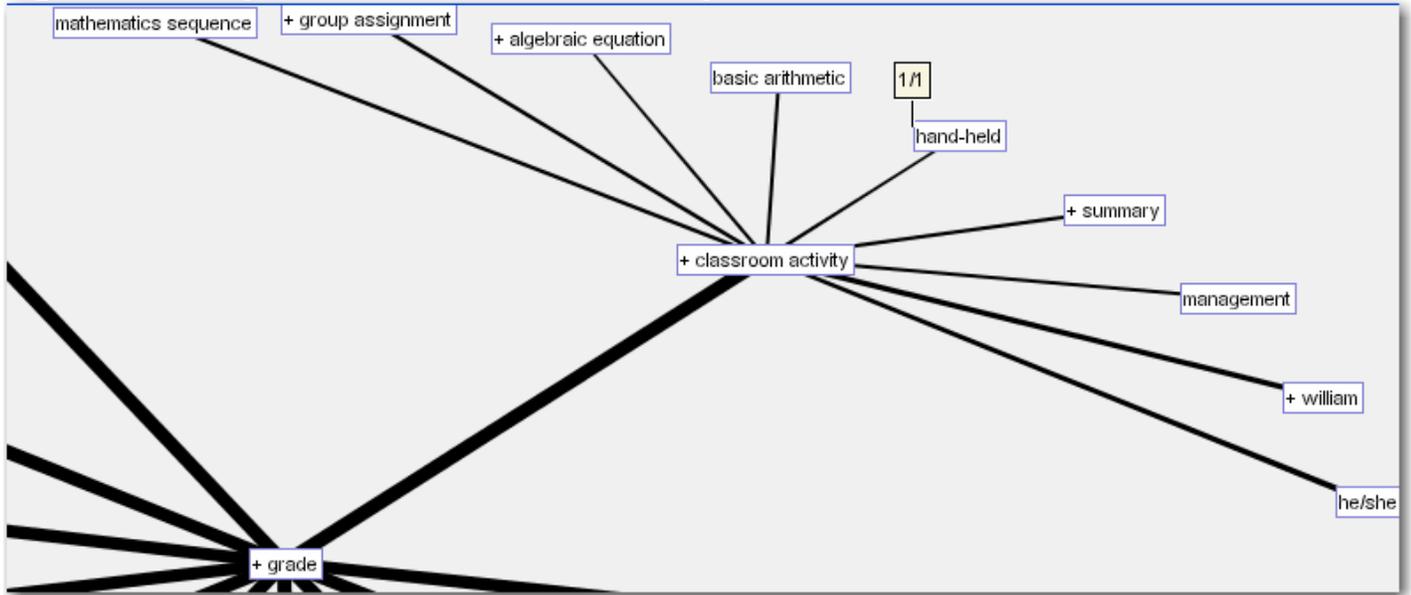**Figure 3. Concept Links to "several quizzes" from "exam"**
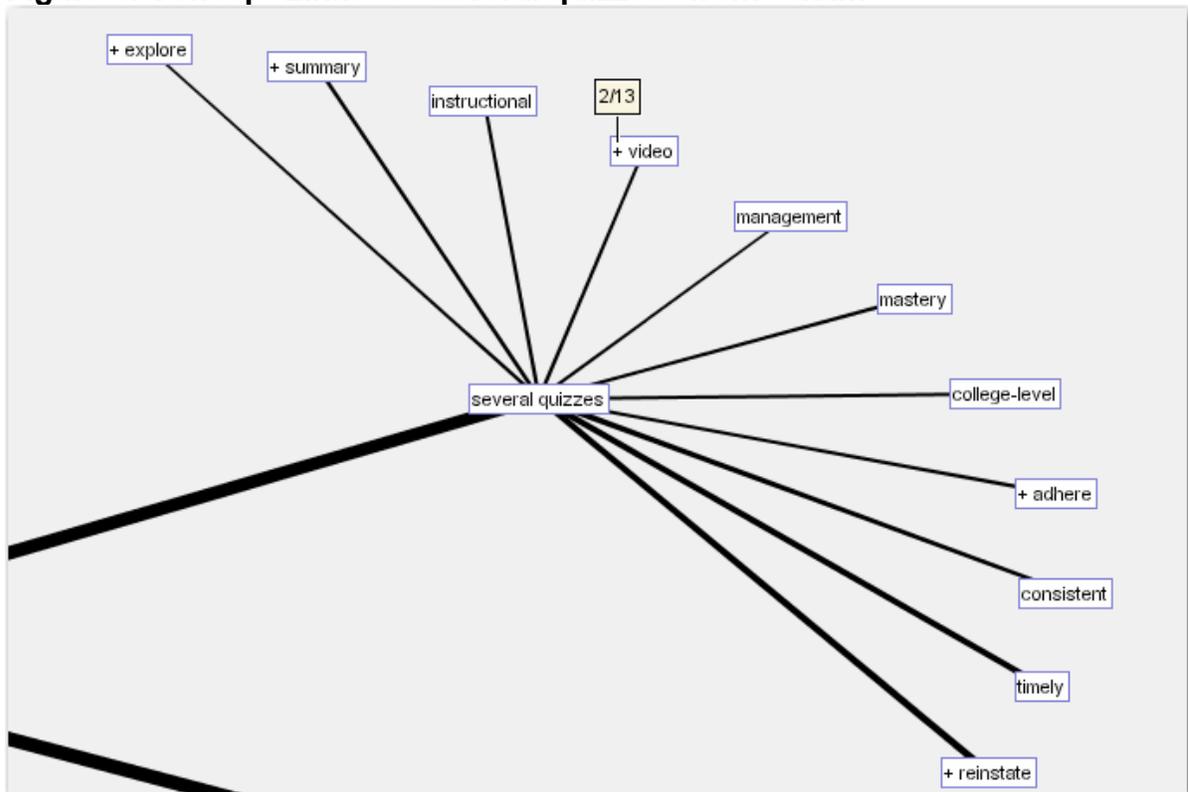
Interestingly enough, the term "grade" is linked to "quizzes," but not to "exam".
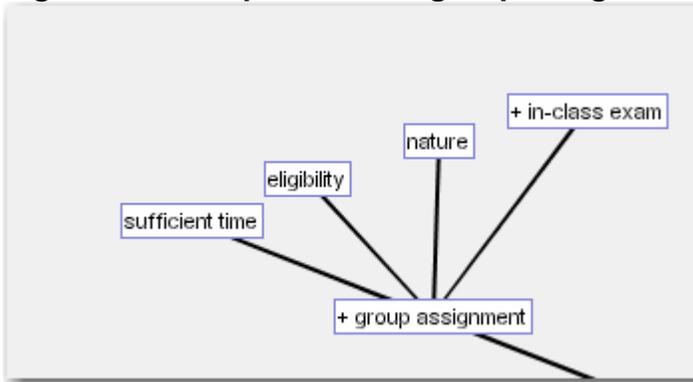
Figure 4. Concept Links to "grade"

There are not many connections to "classroom activity". However, some do deal with "group assignment" and "hand-held", indicating in class work with a calculator.

Figure 5. Concept Links for "classroom activity"

One of the primary connections to "group assignment" is "in-class exam".

Figure 6. Concept Links to "group assignment"

While a handful of course syllabi discuss group work and "hands on" activities, most follow the traditional lecture>practice>exam format. Mathematics instruction remains very traditional.

PhD Requirements

When we examined the terms related to PhD degree requirements, it was found that all but two programs still relied on the model that required qualifying examinations prior to initiating a program of research. A total of 100 documents were examined, with 78 related to PhD degree requirements. The term, "examination," was listed 916 times in the documents. The term, "research" appeared only 454 times. Concept links are used to examine the relationship between examinations and research more closely. Figure 7 gives the clusters for degree requirements. Cluster 1 with 42 documents concentrates on the examinations and other requirements. None of the clusters feature the term, "research," the most prominently.

Figure 7. Clusters of Degree Requirements

| #	Descriptive Terms	Freq	Percentage
1	+ examination, + department, + master, + language, + student, + requirement, + degree, + credit, + 's, + hour	42	0.531645569...
2	+ select, + sequence, + make, + write, + expect, + member, + require, + satisfy, + master, + area	9	0.113924050...
3	+ meet, + follow, + field, + make, + apply, + science, + topic, + application, + write, + have	16	0.202531645...
4	+ hold, + professor, + component, + enroll, + part, + form, + make, + meet, + prepare, + candidate	6	0.075949367...
5	+ contact, + allow, + offer, + include, + prepare, + approve, + apply, + complete, + department, + area	6	0.075949367...

Figure 8 shows the concept links for "examination"; Figure 9 gives the links for "qualifying". Note that the connections are fairly circular, featuring "oral examination", "qualifying", and "qualifying exam". These links clearly demonstrate that exams are the cornerstone of the PhD program.
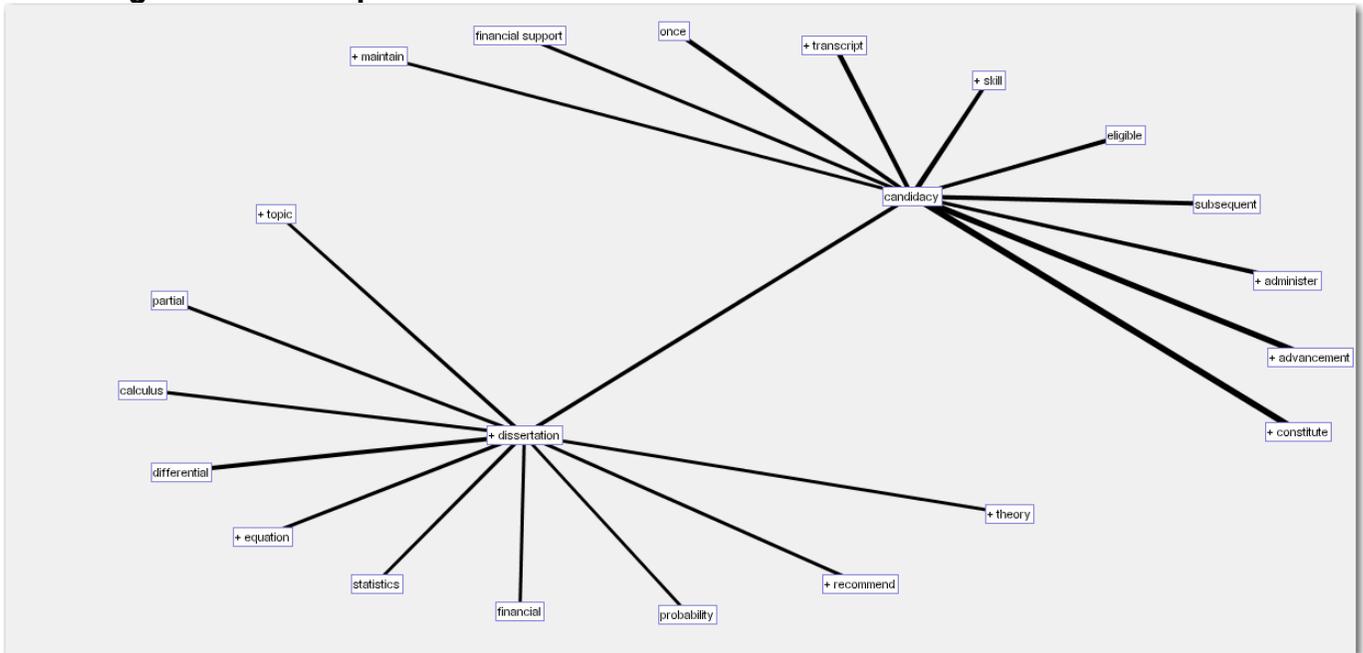
Figure 8. Concept Links for Examination
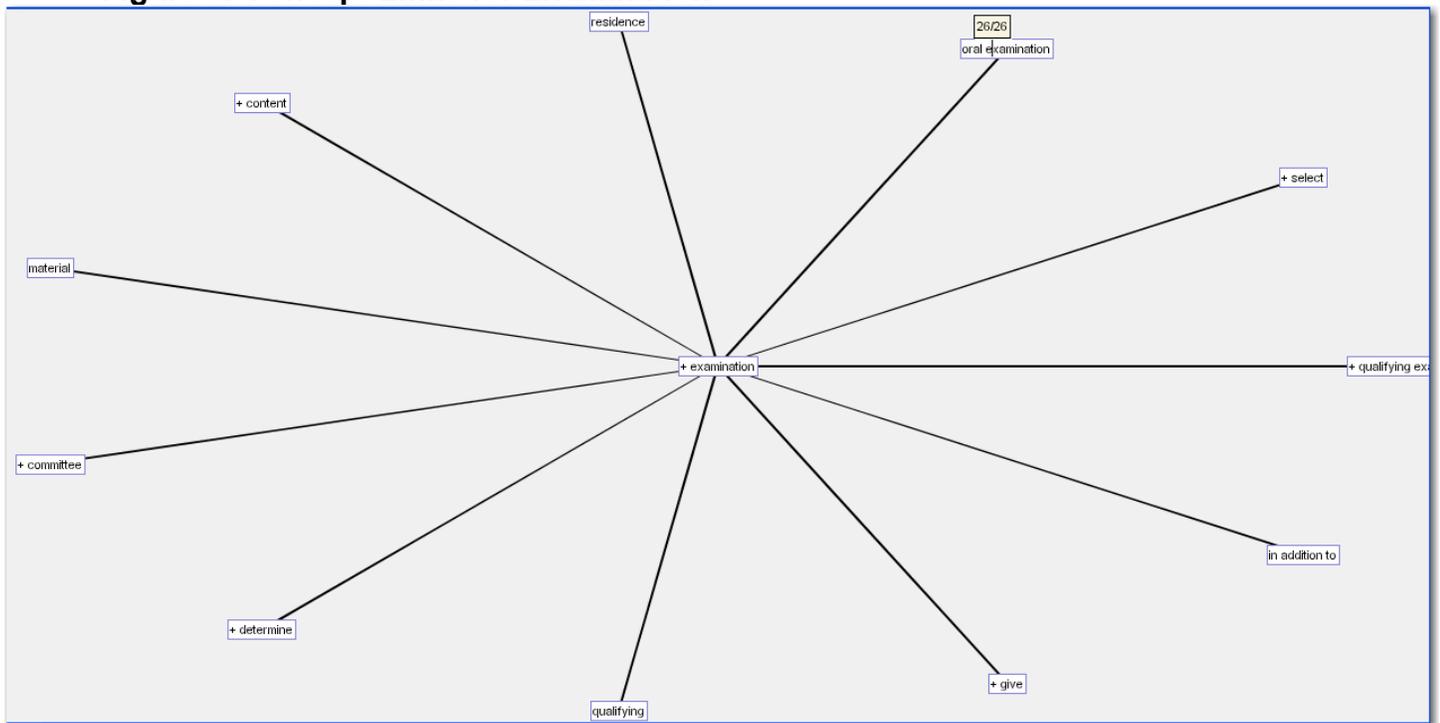
One of the links is to "dissertation," indicating that "dissertation" often follows "qualify". Two of the links, "pure and applied mathematics" and "mathematical foundation" are to examination topics (Figure 9). Figure 10 shows the links to "dissertation".

Figure 9. Concept Links to Qualify
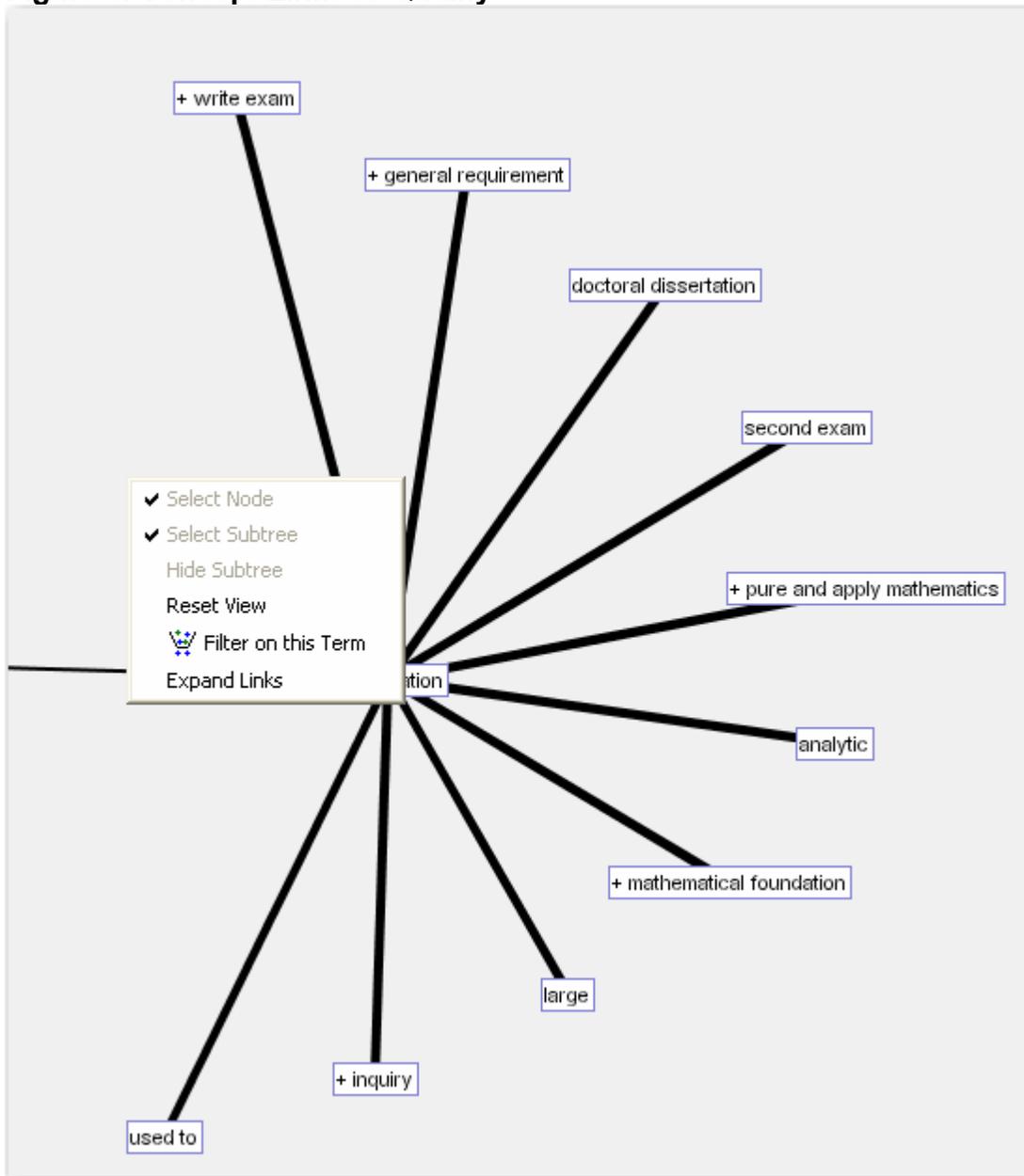
Most of the connections to "dissertation" are for general topics of research, such as "statistics" and "differential equations". As a link to degree requirements, "dissertation" is linked to "candidacy". Most degree requirements do include an advancement to candidacy prior to graduation. Figure 11 gets to the heart of the PhD degree, research.

Figure 10. Concept Links to Dissertation

The primary link from "researcher" is to "candidacy". It indicates that there are few connections from "research", making it the pinnacle of the PhD program, but not a strong part of the program until examinations are completed.

Figure 11. Concept Links for Research
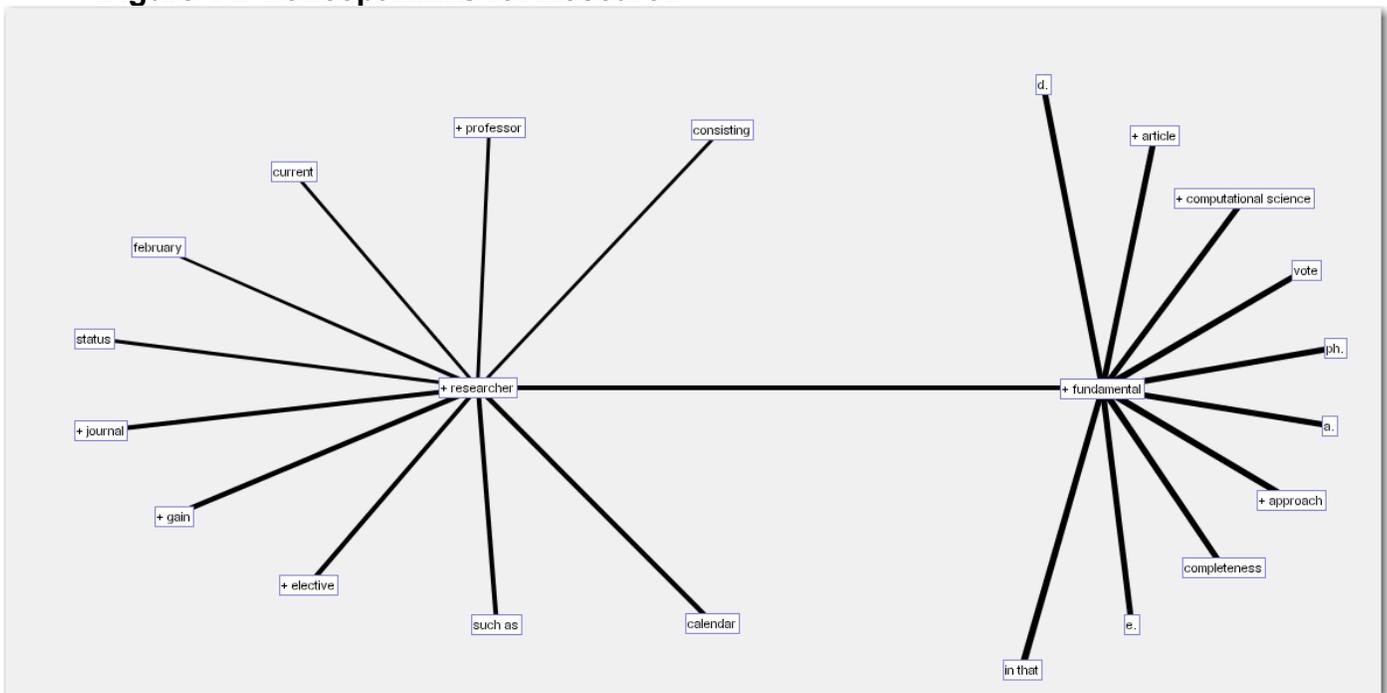
Conclusion

The web and text mining techniques greatly simplify the ability of researchers to examine source documents so that programs and requirements can be compared across institutions in a quantitative fashion. Text mining also allows us to drill down into document details for close examination.