UNIVERSITY OF KENTUCKY
Research and Graduate Studies

The Graduate School
Patterson Office Tower
Lexington, Kentucky 40506-0027
FAX: (606) 323-1928

November 19, 1998

Dr. Raphael Finkel
Department of Computer Science
305 Mining Laboratory
UK Campus 0046

raphael@cs.engr.uky.edu

Dear Raphael:

Your proposal to offer a sequence of courses leading to a Graduate Certificate in Informatics was discussed at the November 11, 1998 meeting of the Graduate Council. The Council agreed that the proposal had substantial merit and that the certificate sequence should have considerable appeal and utility for graduate students in various programs. The Council recommended that I approve this Graduate Certificate.

I approve the **Graduate Certificate in Informatics,** contingent on the following understanding:

1. It is confined to a 15 credit hour curriculum as stated in your proposal of September 29, 1998.

2. If new specializations are added beyond the initial three (humanities, biology, and health care), each specialization will be submitted for Graduate Council review.

3. It is available only to students already enrolled and in good standing in a graduate program at the University of Kentucky.

4. Throughout the proposal the prefix "INF" is used to designate new courses. New courses submitted as part of this proposal must still be approved through the normal academic channels. Use of the INF prefix must also be approved by the Registrar's office at the same time that any new courses designated with this prefix are routed through the required academic councils. In the meantime you could offer the new courses under different sections of GS 600 (Special Topical Graduate Course), which can be repeated for a second semester.

5. You, or some other tenured faculty member associated with the Certificate, will maintain the necessary records to document which students have successfully completed the Certificate sequence.

6. A Certificate signed by the Dean of the Graduate School and the Certificate Coordinator will be awarded to students who have completed the Certificate's requirements as documented in #5 above.

I look forward to learning about the progress of this Certificate sequence.

Best regards,

Michael T. Nietzel, Dean
The Graduate School

MTN:mdc

Certificate in Informatics

29 September 1998

From: Raphael Finkel <raphael@cs.uky.edu>
To: Dean Mike Nietzel, Graduate School

1. Introduction

We propose a **graduate-level certificate curriculum for informatics.**

Computer-related technologies are becoming ubiquitous in all areas of science, engineering, medicine, agriculture, education and humanities. Research and professional success in these disciplines often depends on creative use of the latest advances in computational technology, by which we mean the broad field that includes computer science, computational science, and informatics. Multidisciplinary research projects involving a substantial component of computational technology are becoming the main venue of expanding the borders of knowledge.

Graduate programs in the US must recognize this quickly growing role of computational technology as well as the importance of multidisciplinary projects. This recognition should be reflected in the graduate curriculum. Our objective in this proposal is to develop a curriculum leading to a graduate certificate in informatics to educate a cadre of researchers and professionals with multidisciplinary backgrounds and with substantial understanding of the principles and applications of computational technology.

This curriculum will train graduate and professional-degree students in the uses of computational and information processing technology in their own fields. The students will be able to use this enrichment to become more productive professionals, to further research in their own areas, and to engage in multidisciplinary research relying on computer and information-processing techniques.

2. Details

2.1. Administration

This program will be administered by the Informatics Committee, which will consist, at first, of Raphael Finkel (Computer Science, committee chair), Mirek Truszczyńsky (Computer Science), Kevin Kiernan (English), Ross Scaife (Classics), Chuck Staben (Biology), Gregory Stump (English), and Terry Turner (Medical Center). The Committee must always include at least one member of the Computer Science Department. We intend that it should always include at least one member of each specialization represented by the courses.

The Committee will be in charge of record keeping, admissions, oversight, and student certification.

2.2. Admission requirements

To be admitted to this curriculum and work towards an Informatics Certificate, students must be associated with the University of Kentucky in one of the following categories: (1) enrolled in a degree program and admitted to the Graduate School, (2) enrolled in a professional-degree program, (3) a Resident in the Medical Center, or (4) a faculty or staff member. Students in categories 2–4 will need to be admitted to the Graduate School (possibly as post-bac students, which requires a baccalaureate degree from an accredited university or college). Admission to the curriculum will be subject to application and approval by the Committee. The Committee may establish grade-point or other admission standards.

Other students, such as undergraduates and Post-Baccalaureate students, can enroll for courses in this curriculum, but they cannot receive a Certificate. If courses in this curriculum are over-subscribed, preference will be given to students admitted to the Certificate curriculum.

2.3. Specializations

The curriculum in informatics will have several **specializations**, including humanities, biology, and health care. Other specializations, such as Fine Arts and Library Science, can be added.

2.4. Certificate Requirements

Students will generally be required to complete three courses (nine credits) with no grade lower than B to complete this curriculum.

INF 401G (Informatics fundamentals: Spring)
Any course in the series INF 5xx (Informatics specialization: Fall)
Course INF 699 (Informatics project: Spring)

We do not anticipate offering certificate courses in the summer. Some specializations may impose a higher requirement (not to exceed five courses, that is 15 hours), by requiring several courses at the INF 5xx level. All courses must all be completed within a span of three years.

These courses are mostly new and do not constitute the core courses of any existing program at the University of Kentucky.

2.5. Division of labor

Each of the courses can be taught by one or more faculty. The CS department will undertake to fully staff INF 401G to be offered once a year and half-staff one INF 5xx and INF 699. Biology will staff INF 520 / BIO 520 yearly. English will staff one INF 5xx yearly. The Medical Center will staff INF 530 yearly.

2.6. Resources

Establishing this curriculum will require new resources in the form of release time, teaching assistants, secretarial support, and a budget for software. We will request these

resources in a separate document.

The Computer Science Department is committed to teaching INF 401G as CS 585 Spring 1999. The English Department is committed to teaching INF 510 Fall 1999. The Medical Center is committed to teaching INF 530 in Fall 1999. Biology already teaches INF 520 / BIO 520 yearly. Dean Nietzel of the Graduate School has promised $1000/year to help with clerical tasks.

3. Course outlines

I. **INF 401G: Informatics fundamentals**

A. **Prerequisite.** Junior undergraduate standing; CS students not allowed

B. **Homework.** Approximately 6 assignments to be completed on the computer. The assignments will not involve creating programs, but they will involve using programs and making minor modification to programs. The assignments will be geared to topics of interest to students in the specializations, although no familiarity with those areas is required.

C. **Teaching.** Team-taught by two faculty members, one in CS, one not.

D. **Expected outcomes**

1. Students will become aware of computer techniques that are of general value in conducting and displaying research in non-computer-science fields.
2. Students will have some hands-on experience using computer-based tools.
3. Students will have seen how simple tools can be written in simple scripting languages.

E. **Syllabus**

1. Fundamentals of computer techniques
   a. Computing environments (Linux, Windows, NT): Basic familiarity
   b. Rudiments of networking and internet: Telnet, FTP, Web browsing, search engines, local pages of interest
   c. Intro to Perl: Reading Perl simple scripts
2. Overview of Computer Science: what are the areas and their tools?
   a. Programming fundamentals
   b. Operating Systems
   c. Database management
   d. Networking
3. Data structures and algorithms: informal, example-driven discussion
   a. Arrays, lists, trees, graphs
   b. Recursion
   c. Sorting and searching
4. Regular expressions (software: Perl)
   a. strings, concatenation, alternation
   b. character classes

c. subpatterns
d. data extraction and string substitution
e. extended example: applying regular expressions to converting a bibliographic data base from ASCII to Refer format.

5. Interactive representation techniques (software: Web browsers, Perl)
   a. HTML markups for forms
   b. CGI scripts for dealing with forms
   c. SGML
   d. Text Encoding Initiative
   e. extended example: interactively searching a bibliographic database.

II. **INF 510: Informatics for the Humanities: Electronic Documents**

A. **Prerequisites.** INF 401G

B. **Homework.** Approximately 6 assignments to be completed on the computer (such as creating and parsing an SGML document or generating a concordance for a particular research purpose). Examples throughout the course will be provided by the instructor, but students will work on the assignments using their own texts coming from their own areas of study and research.

C. **Teaching.** Possibly team-taught by two faculty members, one in CS, one not.

D. **Expected outcomes.**
   1. Students will become aware of computer techniques that are of specific value in conducting and displaying research in the humanities, particularly research involving searching online data.
   2. Students will have hands-on experience using computer-based tools that embody these techniques.
   3. Students will have seen how such tools are written.

E. **Syllabus**
   1. **Creation of documents**
      a. Representing cultural objects in digital form (image and text digitization, audio and video digitization; issues of preservation and dissemination)
      b. Image digitization and encoding formats (software: Photoshop)
      c. text digitization (electronic texts in public domain, OCR and transcription)
      d. Document analysis, procedural and descriptive markup, different encoding systems
      e. SGML and TEI (software: Author/Editor, Emacs): Basics of SGML, DTD (example 'TEI lite'), metadata, creating a TEI header, creating an SGML document.
   2. **Analysis of documents**
      a. Text databases (software: Awk, Perl)
      b. Single-file data with one entry per line

c. Simple statistics and manipulation with Awk
d. Multiple related files linked by unique identifiers
e. Hierarchically-organized data in subdirectories of files
f. Extended example: poetry database
g. Searching image databases and image content
h. Databases for qualitative analysis (software: NUDIST)
i. Bibliographic databases (software: htgrep)
j. Text analysis (software: Tact, TactWeb, Wordsmith Tools): concordances, vocabulary and style analysis, authorship studies, metrics
k. Text analysis (software: Perl): word counts, tables of synonyms, concordances: definition, generation and searching, automatic taxonomic
l. Categorization, extended example: themes in Shakespearean sonnets
m. Image processing

3. **Delivery of documents**
   a. Electronic publishing formats (HTML, PDF, SGML) and media (Web, CD-ROM, DVD)
   b. Open standards (SGML, HTML, JPEG, TIFF) and proprietary formats
   c. SGML browser (Panorama Pro, Dynatext)
   d. Converting from one format to another (SGML to HTML), delivering SGML on the Web, XML
   e. Interactive presentation techniques (software: Tcl/Tk): display of frames, buttons, text areas, labels, scrollbars; binding user actions to program operations; collecting user input and producing visible output
   f. Extended example: interactively searching a bibliographic database
   g. Project management (time, costs, planning, documentation, stages of the project and their challenges) and copyright.

III. **INF 511: Informatics for the Humanities: Language**

A. **Prerequisites.** INF 401G

B. **Homework.** Approximately 6 assignments to be completed on the computer The assignments will involve creating only simple programs. Mostly, they will involve using programs and making minor modification to programs. The assignments will be geared to topics of interest to students in the Humanities.

C. **Teaching.** Possibly team-taught by two faculty members, one in CS, one not.
   1. Languages, formal and natural (software: Perl)
      a. regular expressions and regular languages
      b. context-free languages
      c. parsing and generation

2. Knowledge representation (software: Prolog, DATR)
    a. production systems, rule-based systems
    b. Prolog
    c. semantic networks
    d. default inheritance hierarchies
    e. DATR
    f. extended example: Latin verb forms
3. Natural-language processing
    a. Fonts for the world's writing systems (Metafont, Unicode)
    b. Software for generating non-English text
    c. Spelling checkers
    d. Augmented transition networks
    e. two-level morphological transducers for morphological parsing
    f. Automatic and computer-assisted translation

IV. INF 520: Informatics for the Life Sciences: Biology

A. **Prerequisites.** INF 401G (or advanced standing in molecular life sciences)

B. **Homework.** Approximately 6 assignments to be completed on the computer The assignments will involve creating only simple programs. Mostly, they will involve using programs and making minor modification to programs. The assignments will be geared to topics of interest to students in Biology.

C. **Teaching.** Chuck Staben

D. **Cross-listing: BIO 520, Bioinformatics.** This course is established in both on-campus and distance-learning modes, and Dr. Staben has an NSF-CCD grant for materials in this area.

E. **Expected outcomes.**
1. Students will become aware of computer techniques that are of specific value in conducting research in biology.
2. Students will have hands-on experience using computer-based tools that embody these techniques.
3. Students will have seen how such tools are written.

F. **Syllabus**
1. Biological structures: function and characterization
    a. Genes
    b. Proteins
2. DNA methods
    a. Sequencing
    b. Fragment assembly
    c. Restriction mapping
    d. Site detection
3. DNA Databases (Entrez)
    a. File conversions

b. Recombinant DNA

4. Genes
   a. Sequence features
   b. ORFs, URFs, codon usage, Consensus Elements
   c. Gene Identification Tools
   d. RNA structure; PCR primer analysis
5. Proteins
   a. Structure basics
   b. 3D viewing
   c. Secondary structure prediction
   d. Membrane proteins; Consensus sites
   e. Protein sequence analyses
6. Alignment and similarity
   a. Nucleotide and protein
   b. Database alignment tools
   c. Multiple sequence alignment
7. Structural inferences
   a. Homology modeling
   b. Experimental uses of alignments
   c. 3D structure modeling
   d. 3D prediction and viewing
8. Evolutionary inference: Phylogeny
   a. Phylogenetic algorithms and interpretation
   b. Phylogenetic analysis programs
   c. Genomics-DNA
   d. Genomics-RNA (transcriptomes)
   e. Genomics-proteins (proteomes)
   f. Genome mapping-traits and linkage
9. Genome databases (AceDB and relational databases)
10. Other topics
    a. Quantitative genetics
    b. Medical Informatics
    c. Specialized databases (usenet)
    d. Museum specimen data entry
    e. Mapping of species distributions based on specimen database records
    f. Geographic Information System and Remote Sensing technologies in biodiversity studies

V. **INF 530: Informatics for Health Care: Health Informatics**

A. **Prerequisites.** INF 401G

B. **Homework.** Approximately 6 assignments to be completed on the computer The assignments will involve creating only simple programs. Mostly, they will involve using programs and making minor modification to programs. The assignments will be geared to topics of interest to students in Health Care.

C. **Expected outcomes.**

1. Students will become aware of computer techniques that are of specific value in the practice of Health Care
2. Students will have hands-on experience using computer-based tools that embody these techniques.
3. Students will have seen how such tools are written.

D. **Syllabus.**

1. **Introduction to Health Informatics**
   a. History of information technology in health care
   b. History of Health Informatics as a discipline
   c. Distinctions between Health Informatics and related disciplines
   d. Health Informatics research issues
   e. Careers in Health Informatics
2. **The Medical Process**
   a. Goals of health care
   b. The medical record: Admit notes, Progress notes, Discharge summary, Other (nursing notes, vital signs, lab reports)
   c. Health Plans: Fee-for-service, Medicaid, Medicare, managed care, HMO, PPO; Gatekeeper physicians, utilization review, preventive care
3. **Medical Records**
   a. History and use of the medical record
   b. Data capture and entry; errors in data
   c. Query and surveillance systems (COSTAR's MQL)
   d. Paper versus computer records
   e. Examples of computer-based systems (COSTAR, RMRS, TMR, STOR)
4. **Types of Data**
   a. Natural language (physician notes)
   b. Coded data (lab report)
   c. Quasi-coded data
   d. Other data (ECG, X-ray)
5. **Vocabularies and Standard Coding**
   a. Nature of standard vocabularies
   b. Desirable properties of a standard medical vocabulary (complete, unambiguous, semantic network)
   c. Medical vocabulary examples (ICD-9, SNOMED, MeSH, UMLS)

6. **Accessing Data in Clinical Databases**
   a. Queries for clinical care versus queries for research
   b. Query language and tools: SQL, medical query language, query by example, natural-language query
7. **Natural-language processing**
   a. Need for coded data for decision support
   b. Problems and issues in coding natural language data
   c. Methods (vocabulary, grammar, parser)
8. **Patient Monitoring**
   a. When monitoring is used and what is monitored (primarily in ICU)
   b. Technology of monitoring and analog-digital conversion
   c. Example of ECG: technology and automatic computer interpretation of data
9. **Medical Images**
   a. Nature of medical images (X-Ray, MRI)
   b. Digital images
   c. Storing, accessing, transmitting, manipulating, and displaying images
10. **Education and Training**
   a. Infrastructure for learning
   b. Indexing the curriculum
   c. Creating content (courseware, animations and video, internet resources, online tests)
   d. Standards for content (HTML, XML)
   e. Administering the curriculum
   f. Web-based continuing education
11. **Telemedicine**
   a. Nature of telemedicine and reasons for use
   b. Applications: Patient care (dermatology consultation); professional education, administrative
   c. Telecommunication tools (telephone, telemetry, videophone, video-conferencing, internet)
   d. Challenges (insurance reimbursement, physician acceptance, costly networks)
12. **Clinical Information systems**
   a. Single, monolithic systems; Multiple systems (lab, radiology, pharmacy, administrative)
   b. Interfacing different systems
   c. Standards for communicating/interfacing systems (HL7)
13. **Probability and Decision theory**
   a. Decision-making heuristics (representativeness, availability, adjustment and anchoring)
   b. Probability theory

c. Measures of usefulness of diagnostic tests (sensitivity, specificity)
d. Bayes' Theorem
e. Utility functions in decision making

14. **Decision Support and Analysis**
    a. History of knowledge-based decision-support systems in Health Informatics
    b. Modes of operation (passive consulting and critiquing, and active advice)
    c. Issues (maintenance and validation of knowledge base, completeness, need for integration with clinical information systems)
    d. Methodologies (production-rule systems such as MYCIN, protocol systems such as ONCOCIN, cognitive models such as Dxplain)
15. **Information Retrieval**
    a. Retrieving information from public databases, medical records, and clinical repositories
    b. Methods for searching natural language sources: Keyword search (MEDLINE using MeSH vocabulary); free text search (looking for words or phrases used in document)
    c. Scoring algorithms (simple word count, vicinity, thesaurus, natural language processing)
    d. Other approaches (factor analysis, neural networks, feedback from users)
    e. User systems: In health care (MEDLINE, Grateful Med); general (AltaVista, Yahoo)

VI. **INF 699: Informatics Project**

A. **Prerequisites.** Any INF 5xx course. Undergraduates may register for this course by special permission only.

B. **Expected outcomes.**
    1. Students will gain a direct understanding of computer techniques that of specific value in conducting and displaying research in their field of concentration.
    2. Students will have hands-on experience using and developing computer-based tools that embody these techniques.

C. **Syllabus**
    1. Design/implementation project for a computational tool of interest in any specialization. Students may work in cross-disciplinary teams. The project must be approved by the Committee.