1. General Information

1a. Submitted by the College of: PUBLIC HEALTH

Date Submitted: 8/11/2014

1b. Department/Division: Dept Of Biostatistics

1c. Contact Person

Name: Andrea Perkins

Email: andrea.perkins@uky.edu

Phone: 218-2021

Responsible Faculty ID (if different from Contact)

Name: Ramakanth Kavuluru

Email: ramakanth.kavuluru@uky.edu

Phone: 218-2246

1d. Requested Effective Date: Semester following approval

1e. Should this course be a UK Core Course?   No

2. Designation and Description of Proposed Course

2a. Will this course also be offered through Distance Learning?:   Yes [4]

2b. Prefix and Number: BMI 733

2c. Full Title: Biomedical Natural Language Processing

2d. Transcript Title:

2e. Cross-listing:

2f. Meeting Patterns

LECTURE: 45

2g. Grading System: Letter (A, B, C, etc.)

2h. Number of credit hours: 3

2i. Is this course repeatable for additional credit?   No

If Yes: Maximum number of credit hours:

If Yes: Will this course allow multiple registrations during the same semester?

2j. Course Description for Bulletin: This course is a technical introduction to the area of biomedical natural language processing (NLP). In the field of biomedical informatics, this focuses on the common steps in extracting information from textual data that arises from biomedical literature and clinical documents. Topics involve n-gram models, tokenization, POS tagging, and parsing.

2k. Prerequisites, if any: MA 123 (or equivalent) or consent of instructor

2l. Supplementary Teaching Component:

3. Will this course taught off campus?   No

   If YES, enter the off campus address:

4. Frequency of Course Offering: Fall,

   Will the course be offered every year?:   Yes

   If No, explain:

5. Are facilities and personnel necessary for the proposed new course available?:   Yes

   If No, explain:

6. What enrollment (per section per semester) may reasonably be expected?:   25

7. Anticipated Student Demand

    Will this course serve students primarily within the degree program?:   No

   Will it be of interest to a significant number of students outside the degree pgm?:   Yes

   If Yes, explain: Natural language is the main medium of spoken and written communication in most areas of study. The core principles will of be of great interest to statisticians, biomedical informaticians, computer scientists, and the applications will be very useful for biomedical researchers, social and political scientists, and communication majors in making sense of and extracting trends from vast amounts of textual data that inundates the digital world every day. Students from Biostatistics, Computer Science, Statistics, Communication, Library and Information Science, Electrical Engineering, and Pharmacy might find this course of interest.

8. Check the category most applicable to this course: Relatively New – Now Being Widely Established,

   If No, explain:

9. Course Relationship to Program(s).

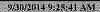   a. Is this course part of a proposed new program?:   No

   If YES, name the proposed new program:

   b. Will this course be a new requirement for ANY program?:   No

   If YES, list affected programs:

10. Information to be Placed on Syllabus.

   a. Is the course 400G or 500?:   No

   b. The syllabus, including course description, student learning outcomes, and grading policies (and 400G-/500-level grading differentiation if applicable, from **10.a** above) are attached: Yes

Distance Learning Form

Instructor Name: Ramakanth Kavuluru

Instructor Email: kavuluru@uky.edu

Internet/Web-based: Yes

Interactive Video: No

Hybrid: No

1.How does this course provide for timely and appropriate interaction between students and faculty and among students? Does the course syllabus conform to University Senate Syllabus Guidelines, specifically the Distance Learning Considerations? Students and faculty will regularly communicate through Blackboard facilitated email and discussion forums. Furthermore, students can contact the faculty member via conventional email or through phone during the office hours. Three live sessions through Adobe Connect will also be conducted to further enhance virtual communication between students and the instructor. Yes, the syllabus conforms to the general senate guidelines and also the distance learning considerations.

2.How do you ensure that the experience for a DL student is comparable to that of a classroom-based student's experience? Aspects to explore: textbooks, course goals, assessment of student learning outcomes, etc. The narrative component part of the lecture is delivered as voice recordings over the slides. However, to render class-room level clarity to the more technical and mathematical aspects, the slides will be designed with expanded intermediate steps with appropriate use of colors and boxed formulae to carefully take the students through the derivations. Assessment, goals, and outcomes will be maintained equal to the class-room offering of the course.

3.How is the integrity of student work ensured? Please speak to aspects such as password-protected course portals, proctors for exams at interactive video sites; academic offense policy; etc. Faculty will be circumspect to monitor plagiarism from online resources and among students; this aspect will be no different from the class-room setting. Password protected Blackboard portal will help retain the integrity and also enforce adherence to submission due dates through time stamping. In addition to stating the academic offense policies in the syllabus, the instructor will emphasize the policies during the beginning of the semester in a recorded session so students are given ample opportunity to familiarize themselves with what is considered plagiarism and what consequences they stand to face when resorting to academic dishonesty.

4.Will offering this course via DL result in at least 25% or at least 50% (based on total credit hours required for completion) of a degree program being offered via any form of DL, as defined above? No

If yes, which percentage, and which program(s)? N/A

5.How are students taking the course via DL assured of equivalent access to student services, similar to that of a student taking the class in a traditional classroom setting? Students can interact with the faculty member through the live sessions or offline office hours via physical visit or through phone calls; other ad hoc means of interactive Web based communication can be perused within the weekly office hour slots. Students will have access to library resources. Students will also be encouraged to take advantage of the DL library resources.

6.How do course requirements ensure that students make appropriate use of learning resources?   Slide shows are necessary but not sufficient in understanding the nuances in this course. The voice recordings are thus indispensable to gain a deeper understanding of the subtleties of the subject. Furthermore, the assignments and exams will also test portions of the text book that are not necessarily covered in detail during the lectures. This further promotes the habit of comprehending abstract concepts from technical manuscripts. Students will also be able to play with interactive online tutorial on SPARQL language that is part of the syllabus that lets them interactively learn a query language.

7.Please explain specifically how access is provided to laboratories, facilities, and equipment appropriate to the course or program.   The assignments will be based on free open source software platforms with proper documentation to install on personal computers. Students will be able to use Portege tool to build their own ontologies. Conventional homework that does not need any programming can be solved using the material (recordings and textbook) and written up using a Word processor.

8.How are students informed of procedures for resolving technical complaints? Does the syllabus list the entities available to offer technical help with the delivery and/or receipt of the course, such as the Information Technology Customer Service Center (http://www.uky.edu/UKIT/)?   Yes. Additional resources for resolving technical issues are stated on the syllabus, including instructions to notify the instructor immediately.

9.Will the course be delivered via services available through the Distance Learning Program (DLP) and the Academic Technology Group (ATL)?   YES

If no, explain how student enrolled in DL courses are able to use the technology employed, as well as how students will be provided with assistance in using said technology.   N/A

10.Does the syllabus contain all the required components?   YES

11.I, the instructor of record, have read and understood all of the university-level statements regarding DL.

Instructor Name: Ramakanth Kavuluru

SIGNATURE|ALHAYS0|Andrea L Perkins|BMI 733 NEW College Review|20140319

SIGNATURE|CHAPPELL|Joseph Chappell|BMI 733 ZCOURSE_NEW Approval Returned to Dept|20140425

SIGNATURE|ZNNIKO0|Roshan N Nikou|BMI 733 NEW Graduate Council Review|20140425

SIGNATURE|ZNNIKO0|Roshan N Nikou|BMI 733 NEW Graduate Council Review|20140428

SIGNATURE|ALHAYS0|Andrea L Perkins|BMI 733 ZCOURSE_NEW Approval Returned to Dept|20140820

SIGNATURE|ALHAYS0|Andrea L Perkins|BMI 733 NEW College Review|20140811

SIGNATURE|ZNNIKO0|Roshan N Nikou|BMI 733 NEW Graduate Council Review|20140926

Courses	Request Tracking

New Course Form

https://myuk.uky.edu/sap/bc/soap/rfc?services=

Open in full window to print or save

Attachments:

Browse... No file selected.

	ID	Attachment
Delete	3272	BMI 733 Syllabus InClass 2-05-14 (Rama).pdf
Delete	3273	BMI 733 Syllabus Online 2-05-14 (Rama).pdf
Delete	3545	CCTS Letters of Support.pdf

First | 1 Last |

Select saved project to retrieve...

(*denotes required fields)

1. General Information

    a. * Submitted by the College of: PUBLIC HEALTH   Submission Date: 8/11/2014

    b. * Department/Division: Dept Of Biostatistics

    c.

       * Contact Person Name: Andrea Perkins   Email: andrea.perkins@uky.edu Phone: 218-2021

       * Responsible Faculty ID (if different from Contact) Ramakanth Kavuluru   Email: ramakanth.kavuluru@uky Phone: 218-2246

    d. * Requested Effective Date: ⦿ Semester following approval OR ◯ Specific Term/Year [1]

    e.

      Should this course be a UK Core Course?   ◯ Yes ⦿ No

      If YES, check the areas that apply:

        ☐ Inquiry – Arts & Creativity   ☐ Composition & Communications - II

        ☐ Inquiry – Humanities   ☐ Quantitative Foundations

        ☐ Inquiry – Nat/Math/Phys Sci   ☐ Statistical Inferential Reasoning

        ☐ Inquiry – Social Sciences   ☐ U.S. Citizenship, Community, Diversity

        ☐ Composition & Communications - I   ☐ Global Dynamics

2. Designation and Description of Proposed Course.

    a. * Will this course also be offered through Distance Learning?   ⦿ Yes [4] ◯ No

    b. * Prefix and Number: BMI 733

    c. * Full Title: Biomedical Natural Language Processing

    d. Transcript Title (if full title is more than 40 characters):

    e. To be Cross-Listed [2] with (Prefix and Number):

    f. * Courses must be described by at least one of the meeting patterns below. Include number of actual contact hours[3] for each meeting pattern type.

45 Lecture	Laboratory[1]	Recitation	Discussion
Indep. Study	Clinical	Colloquium	Practicum
Research	Residency	Seminar	Studio
Other	If Other, Please explain:		

    g. * Identify a grading system:

      ⦿ Letter (A, B, C, etc.)

      ◯ Pass/Fail

      ◯ Medicine Numeric Grade (Non-medical students will receive a letter grade)

      ◯ Graduate School Grade Scale

    h. * Number of credits: 3

    i. * Is this course repeatable for additional credit?   ◯ Yes ⦿ No

      If YES: Maximum number of credit hours:

      If YES: Will this course allow multiple registrations during the same semester? ◯ Yes ◯ No

    j. * Course Description for Bulletin:

      This course is a technical introduction to the area of biomedical natural language processing (NLP). In the field of biomedical informatics, this focuses on the common steps in extracting information from textual data that arises from biomedical literature and clinical documents. Topics involve n-gram models, tokenization, POS tagging, and parsing.

k. Prerequisites, if any:

MA 123 (or equivalent) or consent of instructor

l. Supplementary teaching component, if any: ○ Community-Based Experience ○ Service Learning ○ Both

3. * Will this course be taught off campus? ○ Yes ● No

If YES, enter the off campus address:

4. Frequency of Course Offering.

a. * Course will be offered (check all that apply): ☑ Fall ☐ Spring ☐ Summer ☐ Winter

b. * Will the course be offered every year? ● Yes ○ No

If No, explain:

5. * Are facilities and personnel necessary for the proposed new course available? ● Yes ○ No

If No, explain:

6. * What enrollment (per section per semester) may reasonably be expected? 25

7. Anticipated Student Demand.

a. * Will this course serve students primarily within the degree program? ○ Yes ● No

b. * Will it be of interest to a significant number of students outside the degree pgm? ● Yes ○ No

If YES, explain:

Natural language is the main medium of spoken and written communication in most areas of study. The core principles will of be of great interest to statisticians, biomedical informaticians, computer scientists, and the applications

8. * Check the category most applicable to this course:

☐ Traditional – Offered in Corresponding Departments at Universities Elsewhere

☑ Relatively New – Now Being Widely Established

☐ Not Yet Found in Many (or Any) Other Universities

9. Course Relationship to Program(s).

a. * Is this course part of a proposed new program? ○ Yes ● No

If YES, name the proposed new program:

b. * Will this course be a new requirement ᵅfor ANY program? ○ Yes ● No

If YES ᵅ, list affected programs::

10. Information to be Placed on Syllabus.

a. * Is the course 400G or 500? ○ Yes ● No

If YES, the *differentiation for undergraduate and graduate students must be included* in the information required in **10.b.** You must include: (i) identification of additional assignments by the graduate students; and/or (ii) establishment of different grading criteria in the course for graduate students. (See SR 3.1.4.)

b. ☑ * The syllabus, including course description, student learning outcomes, and grading policies (and 400G-/500-level grading differentiation if applicable, from **10.a** above) are attached.

Distance Learning Form

This form must accompany every submission of a new/change course form that requests distance learning delivery. This form may be required when changing a course already approved for DL delivery. **All fields are required!**

Introduction/Definition: For the purposes of the Commission on Colleges Southern Association of Colleges and Schools accreditation review, *distance learning* is defined as a formal educational process in which the majority of the instruction (interaction between students and instructors and among students) in a course occurs when students and instructors are not in the same place. Instruction may be synchronous or asynchronous. A distance learning (DL) course may employ correspondence study, or audio, video, or computer technologies.

A number of specific requirements are listed for DL courses. **The *department* proposing the change in delivery method is responsible for ensuring that the requirements below are satisfied at the individual course level.** It is the responsibility of the instructor to have read and understood the university-level assurances regarding an equivalent experience for students utilizing DL (available at http://www.uky.edu/USC/New/forms.htm).

Course Number and Prefix:	BMI 733	Date:	3/19/2014
Instructor Name:	Ramakanth Kavuluru	Instructor Email:	kavuluru@uky.edu

Check the method below that best reflects how the majority of the course content will be delivered.

Internet/Web-based ☑   Interactive Video ☐   Hybrid ☐

Curriculum and Instruction

1. How does this course provide for timely and appropriate interaction between students and faculty and among students? Does the course syllabus conform to University Senate Syllabus Guidelines, specifically the Distance Learning Considerations?

   Students and faculty will regularly communicate through Blackboard facilitated email and discussion forums. Furthermore, students can contact the faculty member via conventional email or through phone during the office hours.

2. How do you ensure that the experience for a DL student is comparable to that of a classroom-based student's experience? Aspects to explore: textbooks, course goals, assessment of student learning outcomes, etc.

   The narrative component part of the lecture is delivered as voice recordings over the slides. However, to render class-room level clarity to the more technical and mathematical aspects, the slides will be designed with expanded

3. How is the integrity of student work ensured? Please speak to aspects such as password-protected course portals, proctors for exams at interactive video sites; academic offense policy; etc.

   Faculty will be circumspect to monitor plagiarism from online resources and among students; this aspect will be no different from the class-room setting. Password protected Blackboard portal will help retain the integrity and also

4. Will offering this course via DL result in at least 25% or at least 50%* (based on total credit hours required for completion) of a degree program being offered via any form of DL, as defined above?

   No

   Which percentage, and which program(s)?
   N/A

   *As a general rule, if approval of a course for DL delivery results in 50% or more of a program being delivered through DL, the effective date of the course's DL delivery will be six months from the date of approval.

5. How are students taking the course via DL assured of equivalent access to student services, similar to that of a student taking the class in a traditional classroom setting?

   Students can interact with the faculty member through the live sessions or offline office hours via physical visit or through phone calls; other ad hoc means of interactive Web based communication can be perused within the weekly office

Library and Learning Resources

6. How do course requirements ensure that students make appropriate use of learning resources?

   Slide shows are necessary but not sufficient in understanding the nuances in this course. The voice recordings are thus indispensable to gain a deeper understanding of the subtleties of the subject. Furthermore, the assignments and

7. Please explain specifically how access is provided to laboratories, facilities, and equipment appropriate to the course or program.

   The assignments will be based on free open source software platforms with proper documentation to install on personal computers. Students will be able to use Portege tool to build their own ontologies. Conventional homework that does

Student Services

8. How are students informed of procedures for resolving technical complaints? Does the syllabus list the entities available to offer technical help with the delivery and/or receipt of the course, such as the Information Technology Customer Service Center (http://www.uky.edu/UKIT/)?

   Yes.
   Additional resources for resolving technical issues are stated on the syllabus, including instructions to notify the

9. Will the course be delivered via services available through the Distance Learning Program (DLP) and the Academic Technology Group (ATL)?

   * Yes
   ○ No

   If no, explain how students enrolled in DL courses are able to use the technology employed, as well as how students will be provided with assistance in using said technology.
   N/A

10. Does the syllabus contain all the required components, below? ☑ Yes

    ○ Instructor's *virtual* office hours, if any.

    ○ The technological requirements for the course.

    ○ Contact information for Distance Learning programs (http://www.uky.edu/DistanceLearning) and Information Technology Customer Service Center (http://www.uky.edu/UKIT/Help/; 859-218-HELP).

    ○ Procedure for resolving technical complaints.

    ○ Preferred method for reaching instructor, e.g. email, phone, text message.

    ○ Maximum timeframe for responding to student communications.

    ○ Language pertaining academic accommodations:

       ▪ "If you have a documented disability that requires academic accommodations in this course, please make your request to the University Disability Resource Center. The Center will require current disability documentation. When accommodations are approved, the Center will provide me with a Letter of Accommodation which details the recommended accommodations. Contact the Disability Resource Center, Jake Karnes, Director at 859-257-2754 or jkarnes@email.uky.edu."

    ○ Specific dates of face-to-face or synchronous class meetings, if any.

    ○ Information on Distance Learning Library Services (http://www.uky.edu/Libraries/DLLS)

       ▪ Carla Cantagallo, DL Librarian

       ▪ Local phone number: 859 257-0500, ext. 2171; long-distance phone number: (800) 828-0439 (option #6)

       ▪ Email: dllservice@email.uky.edu

       ▪ DL Interlibrary Loan Service: http://www.uky.edu/Libraries/libpage.php?lweb_id=253&llib_id=16

11. I, the instructor of record, have read and understood all of the university-level statements regarding DL.

    Instructor Name:
    Ramakanth Kavuluru

Abbreviations: DLP = Distance Learning Programs ATG = Academic Technology Group Customer Service Center = 859-218-HELP (http://www.uky.edu/UKIT/Help)

Revised 8/09

---

[1] Courses are typically made effective for the semester following approval. No course will be made effective until all approvals are received.

[2] The chair of the cross-listing department must sign off on the Signature Routing Log.

[3] In general, undergraduate courses are developed on the principle that one semester hour of credit represents one hour of classroom meeting per week for a semester, exclusive of any laboratory meeting. Laboratory meeting, generally, represents at least two hours per week for a semester for one credit hour. (from SR 5.2.1)

[4] You must also submit the Distance Learning Form in order for the proposed course to be considered for DL delivery.

[5] In order to change a program, a program change form must also be submitted.

Rev 8/09

Submit as New Proposal   Save Current Changes

<u>**Course Syllabus**</u>
BMI 733: Biomedical Natural Language Processing
(Distance Learning Offering through Blackboard)

Classroom and Meeting Hours:

- Online course materials posting date: Since this is a distance learning course, new materials are made available every Monday by midnight through Blackboard

- Online live sessions: During the semester, three one hour online meetings will be conducted via UKY Adobe Connect facility during which students can ask questions about the course material or assessment (homework, assignments etc.). This is in addition to email and phone.

- Face to face meetings are not scheduled for this course. Course materials (slides, recordings) can be viewed offline and all assessment is also conducted through Blackboard submissions.

<u>Contact information</u>

o Instructor: Dr. Ramakanth Kavuluru, Asst. Professor, Biostatistics
  - 230E MDS Building, 725 Rose Street

o Telephone: 859-218-2246

o E-mail: rvkavu2@uky.edu (preferred way of communication, maximum response time frame is 48 hours)

o Office Hours: 2 hours will be maintained per week during which phone calls will be taken in additional to regular email communication.

<u>Course description</u>

This course is a technical introduction to the area of biomedical natural language processing (NLP). In the broad field of biomedical informatics, this focuses on the common steps in extracting information from textual data that arises from biomedical literature and clinical documents. The first part of the course will focus on the basic steps of a language-processing pipeline including tokenization, normalization, part of speech tagging, chunking, and full parsing focusing on both theoretical foundations and practical approaches. Students will then have the opportunity to learn about how biomedical knowledge bases like the UMLS Metathesaurus and NLM tools like Metamap are useful in language processing. Basics of the Python programming language will be discussed in the initial sessions to be later used for a few programming assignments. High-level tasks in language processing like named entity recognition, relation extraction, and sentiment analysis will also be introduced.

Course rationale:

In Biomedicine, unstructured text arises in the form of scientific articles that report new results and clinical narratives that form an important component of medical records. Large-scale manual analysis of textual data is impractical and unsuitable to extract information and elicit important trends. However, text is still the main medium of communication, and often important pieces of information recorded in textual narratives do not make into structured sources that are maintained for specific purposes such as patient billing. So for clinical phenotyping, knowledge discovery, and other secondary usages of electronic medical records and literature, computational approaches to processing textual data are important. With the advent and popularity of online social networks, language processing is also useful in analyzing product markets through opinion mining and community formation and information diffusion for campaigns and causes. Thus, this course will open new opportunities in exploiting the vast amounts of textual data to students in the fields such as medicine, public health, computer science, electrical engineering, information and library science, political science, and social media and communications. Besides providing the skills to perform content analysis for research purposes, it will also help students in their career prospects with many companies, in both IT and healthcare domains, hiring graduates with experience in NLP.

Course prerequisites
- MA 123 (or equivalent) and CS115 (or equivalent) or consent of the instructor

Course objectives
1. Define the building blocks of natural languages
2. Identify and use different components of an NLP pipeline
3. Design and employ NLP pipelines to solve biomedical informatics problems
4. Summarize different biomedical knowledge-bases and their applications in NLP
5. Discuss the theoretical framework of basic NLP tasks

Public Health Competency attainment

Your attainment after completing BMI 733 will be at least the following and perhaps more, depending on the other courses in which you have enrolled.

MPH Concentration-Specific Competencies

- Apply the basic concepts of probability, random variation, and commonly used probability distributions.
- Apply and interpret common univariate and multivariate statistical methods for inference.

Competencies for the PhD in Epidemiology and Biostatistics

- Draw appropriate inferences from data.
- Demonstrate proficiency in using computing tools commonly encountered in epidemiology and biostatistics.
- Demonstrate an understanding of concepts of probability and statistical inference as they apply to problems in public health.

Public Health Competencies for Biomedical Language Processing

http://www.phf.org/resourcestools/pages/core_public_health_competencies.aspx
- Uses information technology to collect, store, and retrieve data
- Demonstrates the use of public health informatics practices and procedures
- Retrieves scientific evidence from a variety of text and electronic sources

Student learning outcomes:
Upon completion of this course, the learner will:
1. Identify various practical issues including synonymy, polysemy, and negation in 'understanding' language
2. Perform basic text processing to extract information from textual documents
3. Describe the n-gram language model and use it for document classification
4. Implement algorithms used for different NLP tasks such as segmentation, tokenization, part-of-speech tagging, and parsing.
5. Use open source tools to build an end-to-end NLP pipeline to solve a real-world problem involving textual data

Textbooks - required
(Readings outside the books will be provided in the form of research papers)

- SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing,
  Computational Linguistics, and Speech Recognition, Second Edition, Jufarsky and Martin, prentice hall, 2008. (Abbreviated JM for schedule)

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008 (Free online copy from the authors: http://nlp.stanford.edu/IR-book/) (Abbreviated MRS for schedule)

Course requirements and learner evaluation
Course grades will be based upon evaluation of the following activities:

120 points - two programming assignments with 60 points for each
1. The first programming assignments deals using Naïve Bayes method to classify biomedical citations as related to cancer or not.
2. The second programming assignment deals with using various features as part of a maximum entropy model to identify named entities in text.

40 points - two review quizzes for 20 points each
1. The first quiz involves computing probabilities of phrases in an n-gram model given a training corpus.
2. The second quiz involves parsing sentences with pencil and paper using CKY and arc-eager methods

120 points - individual project that employs NLP tasks such as tokenization, POS tagging, parsing in a real world application (20 points for proposal, 70 points for implementation report, and 30 points for presentation)

120 points - final exam

Points grading scheme:

400-360=A
359-320=B
319-280=C
0-279=E

Submission of Assignments
Assignments should be submitted via Blackboard by 11:59PM, EST on the day they are due unless otherwise specified.

Technology Requirements
Since this is a distance-learning course, students are required to have stable access to a computer and hence encourage students to own a personal computer (laptop/desktop is okay but tablets are not sufficient) according to these requirements:
http://www.uky.edu/ukit/hardwareguide
Furthermore, to fully participate in the distance learning experience students are also required to install the following software and acquire useful hardware:
http://www.uky.edu/DistanceLearning/current/technology/techReqs.html

Technical Issues
Students should contact the instructor (via email) and the UKIT help desk in case they are unable to access the materials (http://www.uky.edu/UKIT, 859-218-HELP) so that issues can be resolved in a timely fashion and homework can be submitted on time.

Contact information for:
TASC   http://www.uky.edu/ukit/atg/tasc
Information Technology Customer Service Center   http://www.uky.edu/ukit/

Distance Learning Library Services
Students are encouraged to take advantage of the library services to find additional online materials or borrow books through inter-library loan (*Carla Cantagallo, DL Librarian (859) 218-1240 Email: dllservice@email.uky.edu*)

Instructor expectations
1. I expect you to review materials from every class session. The components are highly interrelated; missing a class will detract from the learning potential of subsequent sessions.
2. I expect you to actively participate in the discussions via blackboard and also interact with me via email for clarifications and course related discussion.
3. I expect you to submit papers using proper English grammar, syntax, and spelling. You are encouraged to use spell check and grammar check prior to submitting your written work. The Writing Laboratory is available to anyone who may need assistance.
4. I expect (and encourage) you to provide honest and timely feedback regarding the content and process of this course throughout the semester.
5. I expect you to share in the responsibility for making this course an enjoyable and beneficial learning experience.
6. Wikipedia *cannot* be used as a cited reference as noted by a co-founder of Wikipedia! You may use Wikipedia to identify appropriate source material. Remember Wikipedia is *not* peer reviewed!

Academic honesty

Academic honesty is highly valued at the University. You must always submit work that represents your original words or ideas. If any words or ideas used in a class assignment submission do not represent your original words or ideas, you must cite all relevant sources and make clear the extent to which such sources were used. Words or ideas that require citation include, but are not limited to, all hard copy or electronic publications, whether copyrighted or not, and all verbal or visual communication when the content of such communication clearly originates from an identifiable sources. Per university policy, students shall not plagiarize, cheat, or falsify or misuse academic records. Students are expected to adhere to University policy on cheating and plagiarism in all courses. The minimum penalty for a first offense is a zero on the assignment on which the offense occurred. If the offense is considered severe or

the student has other academic offenses on their record, more serious penalties, up to suspension from the university, may be imposed. All incidents of cheating and plagiarism are taken very seriously at the University of Kentucky, and there are specific policies and procedures in place to prosecute them. See S.R. 6.3.0 (PDF) for the exact Senate Rules regarding academic offenses.

Accommodations

If you have a documented disability that requires academic accommodations, please see me as soon as possible during scheduled office hours. In order to receive accommodations in this course, submit to me a Letter of Accommodation from the Disability Resource Center). If you have not already done so, please register with the Disability Resource Center for coordination of campus disability services available to students with disabilities. Contact Jake Karnes via email at jkarnes@email.uky.edu or by telephone 859-257-2754. You may also visit the DRC website for information on how to register for services as a student with a disability: http://www.uky.edu/StudentAffairs/DisabilityResourceCenter/

Religious Observances

Students will be given the opportunity to make up work (typically, exams or assignments) when students notify their instructor that religious observances prevent the student from completing assignments according to deadlines stated in this syllabus. Students must notify the course instructor **at least two weeks prior to such an absence** and propose how to make up the missed academic work.

Inclement weather

The University of Kentucky has a detailed policy for decisions to close in inclement weather. The snow policy is described in detail at http://www.uky.edu/PR/News/severe_weather.htm or you can call (859) 257-1754.

Late work policy

Late home works will not be accepted unless arranged in case of extenuating circumstances

Excused absences policy

Attendance, excused absences and make-up opportunities for this course will conform to the course policies established by the Office of Academic Ombud Services as found at www.uky.edu/Ombud/policies.php

Course schedule and topics with text book chapters for reading assignments (tentative)

Week	Topic	Readings/Assignments
Week 1	Introduction to biomedical text mining tasks including NER and relation extraction, review of linguistic essentials (parts of speech, phrase structure) (Live class through Adobe Connect)	JAMIA NLP intro paper Text Mining for EHRs paper JM Chapter 12
Week 2	Basic text processing in Unix like operating systems, regular expressions	JM Chapter 2
Week 3	Introduction to probabilities, n-gram language models, perplexity, smoothing techniques,	JM Chapter 4
Week 4	n-gram model interpolation and backoff, Naïve Bayes algorithm for text classification	MRS Chapter 13 (Quiz1 due – ngram modeling)
Week 5	Introduction to named entity recognition, relation extraction, and other forms of information extraction	Custom slides (Project proposals due)
Week 6	Conditional vs generative models, maximum entropy models for named entity recognition	JM Chapter 6.6-6.8 (Text classification assignment due)
Week 7	Part of speech tagging using maxent models, relation extraction (supervised, distant supervision)	JM Chapter 5
Week 8	Intro to parsing, PCFGs, CNF, CKY algorithm (Live class through Adobe Connect)	JM Chapter 13, Chapter 14.1-14.3
Week 9	Issues with PCFGs, Lexicalized PCFG, splitting non-terminals	JM Chapter 14.4-14.7 (Named entity recognition assignment due)
Week 10	Dependency parsing, arc-eager parser, Malt parser, relation extraction through dependency structure	Custom slides
Week 11	Biomedical NER (MetaMap) and Relation Extraction (SemRep)	NLM MetaMap JAMIA paper, MetaMap reference paper
Week 12	Lexical semantics, synonymy/homonymy/polysemy, word sense disambiguation	JM Chapter 19 Quiz 2 due: parsing and POS tagging
Week 13	Word similarity, distributional methods (LSA, RI), knowledge based methods (Wu & Palmer, Lin)	JM Chapter 20 and custom slides
Week 14	Inverted index, postings lists, term-document matrices, tf-idf weighting, vector space model	MRS Chapters 1,2,6, and 8 (Project final report and presentations due)
Week 15	Project presentations through (Live class through Adobe Connect)	
Week 16	Review for final exam (Live class through Adobe Connect)	Final exam question paper uploaded to Black Board by Thursday 11:59 PM of the week before finals week
Final exam take home due by 11:59PM via Blackboard on the Tuesday of the finals week		