WKU A LEADING AMERICAN UNIVERSITY WITH INTERNATIONAL REACH

UNIVERSITY OF LOUISVILLE

J.B. SPEED SCHOOL OF ENGINEERING
Computer Engineering and Computer Science

JOINT RESEARCH PROPOSAL

BETWEEN

The Office of Distance Learning
Western Kentucky University
AND
Knowledge Discovery and Web Mining Lab
Dept. of Computer Engineering and Computer Science
J.B. Speed College of Engineering
University of Louisville

NSF

A Joint Research started on January, 2010

WKU A Leading American University with International Reach

University of Louisville

J.B. Speed School of Engineering
Computer Engineering and Computer Science

- **Dr. Leyla Zhuhadar:**

Research Scientist at the Office of Distance Learning, Western Kentucky University.

- **Dr. Olfa Nasraoui:**

Associate Professor and Endowed Chair of E-Commerce, Director of Knowledge Discovery and Web Mining Lab, University of Louisville.

- **Dr. Robert Wyatt:**

Associate Professor, Director of the Office of Distance Learning, Western Kentucky University.

- **Dr. Adel S. Elmaghraby:**

Professor and Department Chair of CECS, Director of Innovative and Emerging Technologies Lab, University of Louisville.

Non-profit Research Agreement

This is a non-profit research project that aims to provide online students and faculty with innovative technologies using Knowledge Discovery and Intelligent Web Applications to improve interdisciplinary learning.

Significance of this Research

The influence of our research on the State of the Art has been witnessed by:

1. As of 2009, the HyperManyMedia search engine has been ranked number 24 on "The Ultimate Guide to Using Open Courseware [1]" (between Cambridge University and Harvard Business). The next five are Princeton, Stanford, Yale, Johns Hopkins, and Boston College. We consider this external recognition as an objective validation of what has been accomplished so far.

[1] http://www.collegedegree.com/library/college-life/50-Open-courseware-writing-classes

Significance of this Research Cont.

The influence of our research on the State of the Art has been witnessed by:

2. Its contribution to over 30 publications in International Conferences (United Kingdom, France, Italy, Spain, Bulgaria, Canada, Mexico, and Netherlands Antilles), Journals, Book Chapters, and two Books published by Springer (in process), please refer to award and publications , section 6.3, for more details.

Significance of this Research Cont.

The influence of our research on the State of the Art has been witnessed by:

3. HyperManyMedia E-learning Platform, as an approach that supports domain specific information retrieval-- E-learning, is the primary contribution to the State of the Art made by this research and the work described in it. This approach is significantly different than those currently used in the development of information retrieval systems. First, it utilizes ontologies as models to provide semantic information. Secondly, this approach uses two different types of ontologies: a global ontology model that represents the whole E-learning domain and a user ontology model that represents the learner profile. Moreover, the implementation of the ontology models is separate from the design and implementation of the information retrieval system. Potentially, the most significant element in the design of this research is the reuse of the same ontology structure in five different facets by extracting or converting the ontology into another structure to fit the purpose of the design.

The Agreement

On January 2010, Dr. Zhuhadar initiated a joint agreement between the Office of Distance Learning at Western Kentucky University and the Knowledge Discovery and Web Mining Lab at University of Louisville. This agreement was supported by the Vice President of Research at Western Kentucky University, Dr. Richard Bowker and the Interim Director of Office of Sponsored Program (OSP), Dr. Steven J. Haggbloom.

The Agreement Cont.

This agreement included the following:

1. Dr. Nasraoui has made contributions in data mining and stream data mining, including clustering using robust statistics and meta-heuristic optimization methods, with applications to web mining, and more recently to scientific data mining in the context of astronomical data sets. She has also contributed toward designing algorithms and evaluation frameworks and metrics for clustering noisy data streams, and made pioneering contributions to the field of mining Web clickstreams and Web personalization. Her research has brought in more than 1.2 Million dollars in funding as a PI, including an NSF CAREER Award. The proposed collaboration overlaps greatly with the scope of her research and her experience in acquiring funding will prove very helpful to this project. Dr. Nasraoui will provide support to WKU research in the following areas:

   - Collaborate with Dr. Zhuhadar in designing a strategic plan to enhance and implement the research.
   - Collaborate with Dr. Zhuhadar in the research from the scientific standpoints in all issues related to designing algorithms and architectural design for HyperManyMedia platform suggested and created by Dr. Zhuhadar.
   - Contribute with ideas to enhance the research.
   - Help in finding the right solicitations from NSF that fit this research and revising the grant proposals written by Dr. Zhuhadar (note: this work starts in phase II, refer to [Phase-II] for more details).

The Agreement Cont.

This agreement included the following:

2. Dr. Elmaghraby, Professor and Department Chair of the Dept. of Computer Engineering and Computer Science and the Director of Innovative and Emerging Technologies Lab agreed to provide a non-profit support to advance the research at the Office of Distance Learning at Western Kentucky University . As part of this support, he agreed that Dr. Leyla Zhuhadar can use the following facilities at University of Louisville:
    - The Knowledge Discovery & Web Mining Lab: This Lab is Directed by Dr. Nasraoui: It conducts research to advance state of the art in the area of Knowledge Discovery in Data sets (KDD), with an emphasis on Data Mining, and in particular Web Mining and Stream Data Mining. The daily activities consist of learning, investigation, design, implementation, testing, and evaluation of efficient algorithms and techniques to solve challenging problems in support of a variety of applications, such as: Web analytics & Web personalization for e-commerce and information retrieval Mining evolving data streams with an emphasis on evolving Web clickstreams Scalable and/or personalized information retrieval in data such as text and astronomical data sets, Web Mining Lab's website [2].
    - CECS Speed School (high-performance computing (HPC) Cluster and Visualization Wall as part of Bioinformatics Lab: A visualization wall with 18 monitors, powered by a cluster of Dell servers. As visualization nodes, U of L is using nine Dell PowerEdge 2950 servers with Intel Xeon 5400 series processors and another PowerEdge 2950 server as the master node. The 18 display units in the wall are Dell UltraSharp 3007WFP wide screen monitors. Dell PowerConnect 6224 Ethernet switches provide the interconnect. The software that manages the cluster is Platform Open Cluster Stack (OCS) sourced through Dell. It gives us a platform to share that information and also to analyze that information as a whole as opposed to individually, HPC WALL 'sWebsite [3].

[2] http://webmining.spd.louisville.edu/index.html
[3] http://louisville.edu/speed/computer/spotlights/cecs-speed-school-hpc-cluster-and-visualization.html

The Agreement Cont.

This agreement included the following:

3. All the work presented and implemented by Dr. Zhuhadar is supervised by the Director of Distance Learning, Dr. Robert Wyatt at Western Kentucky University .

The Agreement Cont.

This agreement included the following:

4. Dr. Zhuhadar's responsibilities are:

- Visiting University of Louisville Labs on a weekly-basis (Friday through Sunday) to collaborate with Dr. Nasraoui and peers of 8 researchers (Doctoral Students). The main goal of these visits is to use High Performance Computers (HPC) and share ideas with colleagues and professors to better enhance the research implemented on the HyperManyMedia Platform at Western Kentucky University, refer to Sections [Phase-I, Phase-II, Phase-III].
- Designing new algorithms for the HyperManyMedia Platform, refer to Sections [Phase-I, Phase-II, Phase-III].
- Implementing those algorithms on a real platform used by the online learner community, HyperManyMedia, refer to Sections [Phase-I, Phase-II, Phase-III].
- Collecting data to better understand the usage of the platform by online students and Faculty, refer to Sections [Phase-I, Phase-II, Phase-III].
- Writing journal articles, proceeding papers, and book chapters related to HyperManyMedia research, refer to Sections [Phase-I, Phase-II, Phase-III].
- Presenting the research in Conferences (note: any of the partners can present the accepted papers), refer to Sections [Phase-I, Phase-II, Phase-III].
- Seek grant opportunities from NSF to advance the research.

Strategic Plan for each Phase of this Research

The research plan evolves over three phases

Strategic Plan for each Phase of this Research Cont.

1. Phase I (January 2010 – 2011)
   Phase I and subsequence commitments for support spans from January 2010 to January 2011:

   - Using University of Louisville Research Labs.
   - Collaborating with University of Louisville research peers.
   - Designing New Algorithms.
   - Implementing and Testing results on the WKU platform HyperManyMedia.
   - Collecting online students/faculty data and feedback.
   - Submitting articles to Journals and Conference Proceedings.
   - WKU is the only entity that provides a stipend to Dr. Zhuhadar for supporting her research ($1000 on a monthly basis). The starting date of this agreement was January 2010 and the commitment made was to continue it until the grant proposal from NSF got approved.

Strategic Plan for each Phase of this Research Cont.

1. Phase II (January 2011 – 2012)
   Phase I and subsequence commitments for support spans from January 2011 to January 2012:

   - Dr. Zhuhadar is responsible for Writing grant proposals by choosing the best solicitations from NSF that fit this research. Also, Dr. Nasraoui is responsible for collaborating and revising all the written grant proposals by Dr. Zhuhadar.

   - Dr. Zhuhadar will also work along with appropriate WKU officials who can best assist in writing the grant/research opportunities.

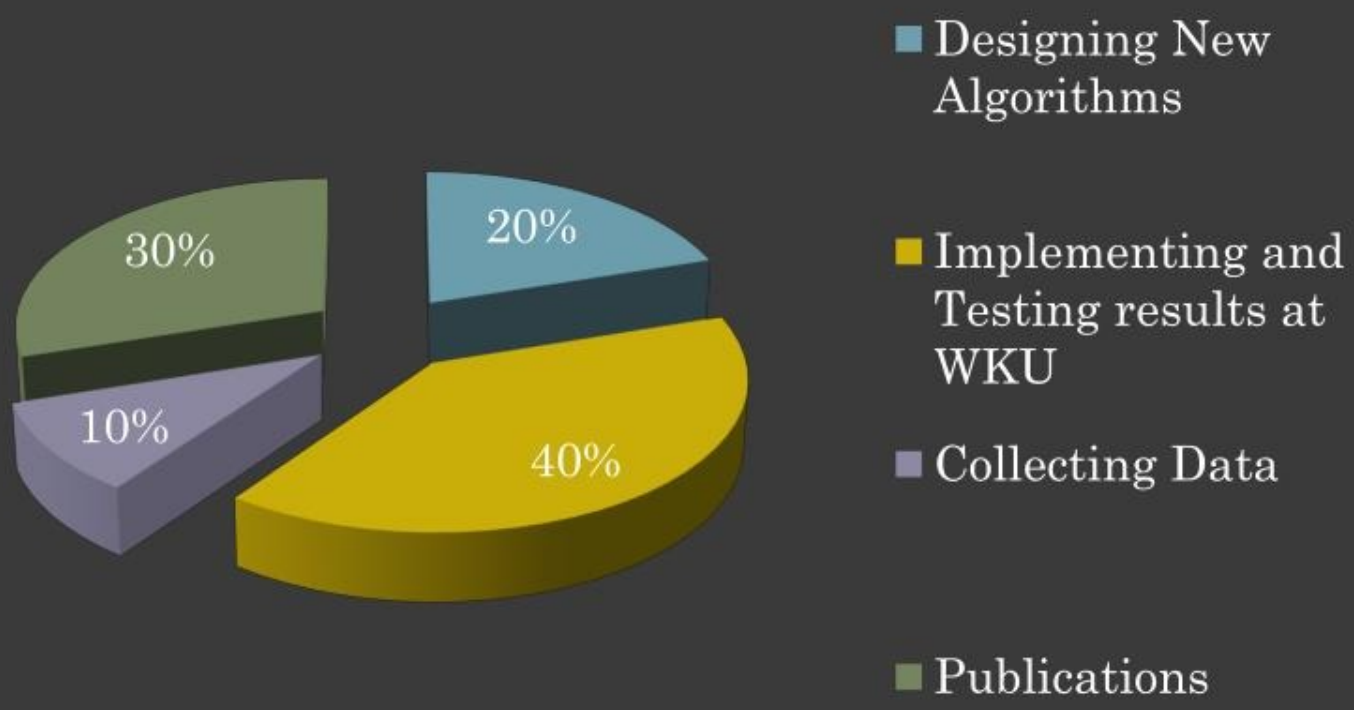
Strategic Plan for each Phase of this Research Cont.

1. Phase III (After 2012)
   Phase III and subsequence commitments for support for support continues after January 2012 :

   - Performing up-to-date innovative research that helps the online student community achieve better learning. The main focus is to develop creative applications that follow the main objectives of S.T.E.M. (Science, Technology, Engineering, and Mathematics). S.T.E.M. is strongly supported by the National Science Foundation and the United States Department of Labor to support and advance the economic role played by a STEM-savvy workforce.

NOTE: It is possible that an external grant supporting the research could be obtained during Phase II. However, it is more likely that a grant could be obtained during Phase III.

PHASE I: JANUARY-SEPTEMBER (2010) TIME SPENT IN EACH OBJECTIVE



PRE-PARTNERING WITH UNIVERSITY OF LOUISVILLE: INDICATORS OF SUCCESSFUL ACCOMPLISHMENTS AT THE OFFICE OF DISTANCE LEARNING, WKU

Growth in Enrollments:

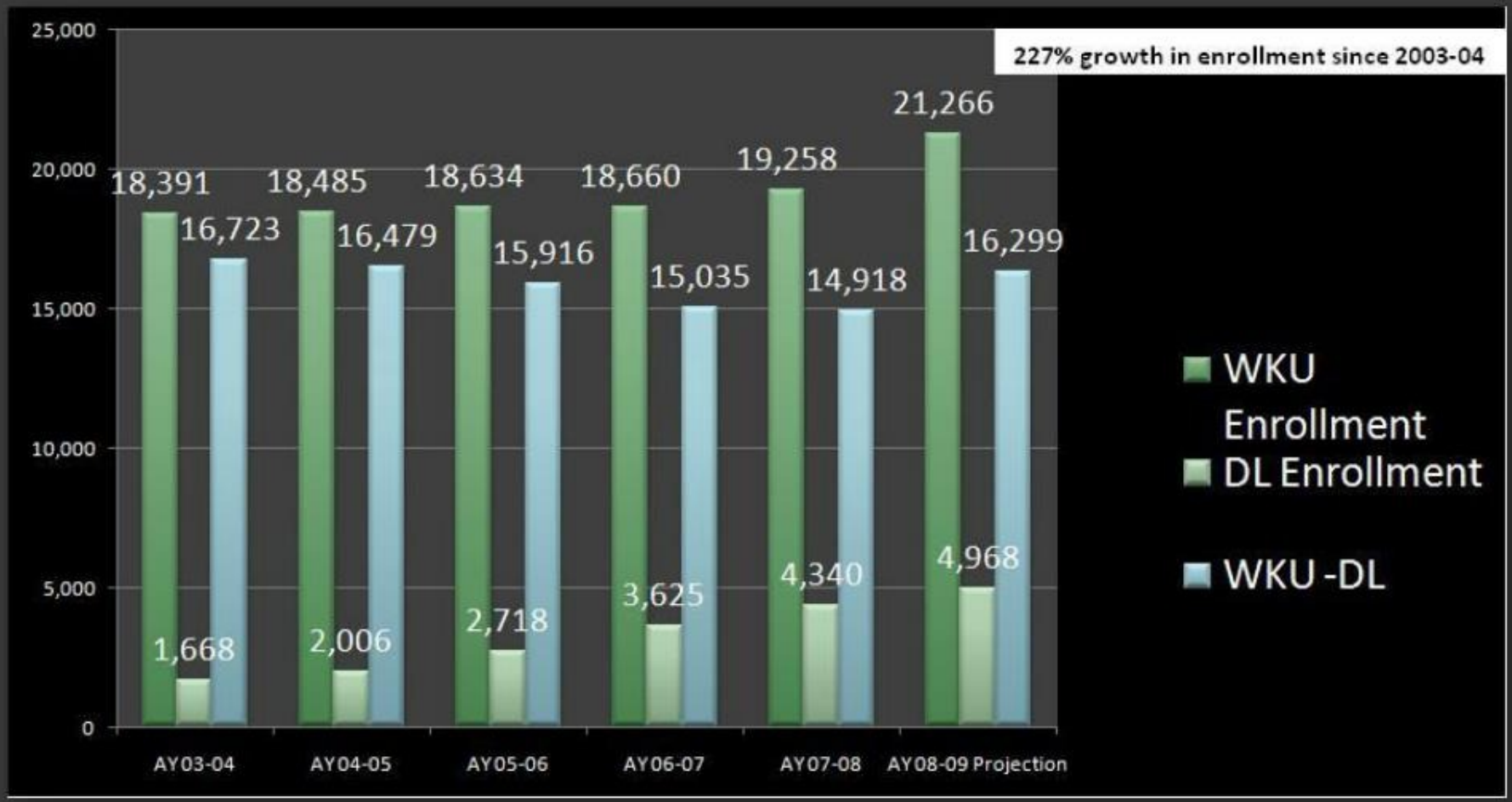
27 programs completely online Last year in 2009

- 1,666 students enrolled online exclusively
- 1,115 graduate
- 551 undergraduate

• Enrollments from

- 103 counties in Kentucky
- 46 states in the U.S.
- 33 countries around the world

PRE-PARTNERING WITH UNIVERSITY OF LOUISVILLE: INDICATORS OF SUCCESSFUL ACCOMPLISHMENTS AT THE OFFICE OF DISTANCE LEARNING, WKU



PRE-PARTNERING WITH UNIVERSITY OF LOUISVILLE: INDICATORS OF SUCCESSFUL ACCOMPLISHMENTS AT THE OFFICE OF DISTANCE LEARNING, WKU

WKU accounts for more than 25% of Distance Learning In Kentucky for both enrollments and headcount

Our online rank in Kentucky

DL Course enrollments	Fall 03	Fall 04	Fall 05	Fall 06	Fall 07	Fall 08
WKU	3,662	3,174	4,109	6,271	7,997	9,440
Morehead State University	1,939	2,549	2,668	4,158	4,961	5,742
Eastern Kentucky University	2,864	3,561	3,817	4,435	4,930	6,367
Murray State University	1,969	1,950	2,205	2,366	2,639	2,883
University of Kentucky	1,603	2,435	2,535	2,967	2,560	3,140
University of Louisville	1,232	1,523	2,214	2,369	2,560	3,265
Northern Kentucky University 3,502		684	904	1,058	1,620	2,271
Kentucky State University	584	869	753	468	1,197	1,043
4-Year Institutions Total	**14,537**	**16,965**	**19,359**	**24,654**	**28,715**	**35,382**

DL Headcount						
WKU	2,410	2,365	2,995	4,120	4,879	5,709
Eastern Kentucky University	2,259	2,766	2,893	3,318	3,679	4,490
Morehead State University	1,446	1,877	1,914	2,650	2,924	3,484
University of Kentucky	1351	1682	1834	2,088	1,683	2,161
University of Louisville	914	1,053	1,472	1,594	1,667	2,053
Murray State University	1,402	1,308	1,420	1,445	1,521	1,717
Northern Kentucky University	554	666	720	1,065	1,373	2,110
Kentucky State University	462	622	510	365	702	699
4-Year Institutions Total	**10,798**	**12,339**	**13,758**	**16,645**	**18,428**	**22,423**

Future Work

Our main goal is to create up-to-date technology that grants us the possibility to have a grant accepted by NSF. We admit that there are many areas in which this research can be extended.

1. In the domain of natural language processing: There are several forms that can be used to extend the multilingual information retrieval system; for example, controlled vocabulary techniques are very well developed, but an automatic construction of the thesaurus is still in its early stages of development. Another area of research that could be beneficial is to consider building the manual thesauri not only based on the controlled vocabulary extracted from the domain ontology as concepts/subconcepts, but by using computational linguistics; in this case, an integration between the thesauri and techniques based on corpus statistics is needed.
2. In the domain of the Semantic Web: "Linked Data" is the right place to extend this research. Linked Data is a project directed by Christian Bizer, Tom Heath and Tim Berners-Lee. The current Web is not structured and the ontology that we designed is considered as a domain specific ontology. It would be more beneficial to use Semantic Web technologies to publish our ontology as structured data on the Web and use the set links between data sources. For example "LinkingOpenData" allows developers to publish existing datasets as Linked Data on the Web and interlink between data sources. In 2009, Linked Data extended into 6.7 billion RDF triples and around 149 million RDF links. Based on this growth, Linked Data might be the future of the Semantic Web.

Finally, we would like to note that our main goal is to help the online community improve its opportunity to learn, share and advance using advanced technology, knowledge discovery and an open source environment!