

## Research Interests

My research interests are broadly in Knowledge Management and Natural Language Processing. In Particular, I am interested in developing Machine Learning and Statistical Modeling techniques for advancing the technologies of Statistical Machine Translation, Information Extraction, Unstructured Data Management and Business Intelligence using unstructured data. I have participated in the research and development of various projects that helped in forming the state of the art in these fields and have been deployed in various solutions to solve customers problems worldwide. I have 5 pending patents, an IEEE journal paper and several papers in the top conferences of the field.

## Education

**Ph.D in Computing, Dublin City University** January, 2009

**M.Sc in Computer and Engineering, Faculty of Engineering, Cairo University** 2001

**B.Sc in Computer and Engineering, Faculty of Engineering, Cairo University with Honors degree** 1996

## Work Experience

### **IBM Cairo Technology Development Center**

Researcher October 2001 : Current  
Working as a Researcher with IBM Watson Research Center, Yorktown Heights - NY, in Natural Language Processing, Machine Translation and Information Extraction.

### **Dublin City University**

PhD Candidate 2005 : 2008  
PhD Candidate at School of Computing with Prof. Andy Way and Dr. Khalil Sima'an, I was working on developing novel techniques for integrating syntax into Phrase-based SMT.

### **IBM Cairo Technology Development Center**

Software Architect August 1999 : September 2001  
Worked as System Architect and Development Team Leader for IBM digital rights management solutions with IBM Forte Lauderdale Lab, Florida.

### **IBM Cairo Technology Development Center**

Research and Development Engineer November 1996 : July 1999  
Worked as Research and Development Engineer in IBM Speech Recognition systems as a member of a multi-national team in EMEA.

## Selected Projects

### **Incremental Dependency Parsing for Statistical Machine Translation**

2007-2008

This project is part of my PhD thesis. I have introduced a novel model of dependency phrase-based SMT which integrates fully incremental dependency parsing into the translation model with the same efficient computational cost assumed in conventional Phrase-based SMT systems. This approach presents an integrated framework for lexicalized, syntactic and semantic machine translation for natural languages and logical languages as well. The work done in my PhD dissertation is considered to be a paradigm shift for enabling syntax and semantic withing state of the art SMT.

### **Syntactic Phrase-based Statistical Machine Translation**

2005-2007

This project is part of my PhD thesis. I have introduced a novel approach for incorporating

syntactic information into state of the art Statistical MT. The approach has provided significant improvement and has been received very positively by the research community. The work has been published in two papers in highly prestigious conferences and an IEEE journal paper.

### **Spoken Language Translation System**

2007

This project is a part of my PhD work. In this work I have introduced a machine translation system for spoken touristic dialogs. The system has been ranked as the “BEST“ machine translation system by human evaluators in an official evaluation campaign (IWSLT, 07). More than ten top universities had participated in the evaluation.

### **Meadan Project**

2006:2007

Meadan is an innovative global dialog blog space that enables dialogs on emerging events across linguistic and cultural boundaries. I have worked with IBM Watson Research team to enhance IBM machine translation system for translating English to Arabic and vice-versa.

### **TALES Project**

2004:2006

TALES is a multimedia mining and translation system that could extract information from websites and broadcast video and audio. I have participated with IBM Watson Research team in creating TALES, developing novel innovative techniques for Information Extraction and advancing the state of the art of this field.

### **InfoMind Project**

2004:2006

InfoMind is a knowledge extraction and representation system. InfoMind provides sophisticated unsupervised information extraction techniques and novel representation and visualization approaches that can represent complex information effectively. The system is being used for providing solutions for knowledge management in the news domain. The system has also been used in the automatic generation of multilingual electronic content from unstructured data for a prototype e-learning system. Two patents have been filed describing this work and two papers in highly prestigious conferences have been published.

### **Semi-Supervised and Unsupervised Machine Learning Techniques**

2005:2006

I have developed novel semi-supervised and unsupervised machine learning techniques that have been successfully deployed in various problems especially for information extraction. The proposed techniques have been published in major conferences and are being a part of the reading lists for machine learning post graduates courses at top universities.

### **iMonitor: Detecting and Predicting Anomalous Situations**

2005-2006

Complex systems require continuous monitoring to keep them running. This requires highly skilled professionals and business experts to detect or predict potential problems in the system, identify their causes, and take appropriate corrective or preventive actions. iMonitor is a solution that integrates Data Mining and Machine Learning Techniques with Common Event Infrastructure for detecting and predicting anomalous situations. This work is being deployed in IBM Business Process Monitoring System. We have filed two patents during this work and it has been selected as one of the most important innovations “**Innovations that Matters**” for IBM at 2007.

### **Automatic Content Extraction (ACE)**

2003-2005

ACE is a NIST evaluation campaign that was focusing on advancing information extraction technologies in Named Entities detection and tracking and relation detection. I have participated in the research and development of various ACE systems. I have developed the Arabic ACE system that had been ranked first in NIST ACE 2003 evaluation.

### **IBM Machine Translation System**

2001-2004

I have participated in the research and development work of IBM Machine Translation System from the early start of this research effort. The performance of IBM system is defining the state

of the art in the field and has been top ranked in yearly NIST evaluation form 2002 till now. I have provided many enhancements for the system such as large translation and language models, Named Entity translation and incorporating syntactic knowledge into SMT system.

### **EMMS: IBMs Electronic Media Management System**

1999-2001

EMMS is a system for providing Digital Rights Management and Protection for video, music, documents and rich media. I was acting as the team leader for the development team, as well as the system Architect for the solution globalization and adding video and documents management and protection. During this project we worked with IBM Fort Lauderdale Lab, FL.

During my master thesis work I have implemented a general purpose Maximum Entropy toolkit using Improved Iterative Scaling for parameter estimation and supported features induction capability. The toolkit, called METK, was totally implemented using C++ STL and had a very good performance. The toolkit had been used in Arabic Topic Classification.

### **Speech Recognition Systems**

1996 : 1999

Acted as member of EMEA speech group focusing on building IBM speech recognition system for both telephony and desktop. The main responsibilities were building and tuning acoustics and language models for IBM ViaVoice telephony and desktop systems. I have participated in the design and implementation of IBM Arabic ViaVoice systems with various releases that had been widely used by end users at that time.

## **Publications**

### **Thesis**

#### **Lexical Syntax for Statistical Machine Translation**

Dublin City University (2009)  
Hany Hassan

### **Journal**

### **Publications**

#### **Syntactically Lexicalized Phrase-Based SMT**

IEEE Transactions on Audio,Speech and Language Processing (2008)  
Hany Hassan, Khalil Simaan and Andy Way

## **Patents Pending**

#### **Detecting and Predicting Anomalous Events**

Filed by IBM 2008  
Method and System for Detecting and Predicting Anomalous Events

#### **Multilingual Browsing Capability**

Filed by IBM 2007  
Method and System for Access of Multilingual Textual Resources using Conceptual Representation Matching

#### **Anomalous Detection**

Filed by IBM 2007  
Method and System for Detecting Anomalous Behavior in Business Process Performance

#### **Knowledge Extraction and Representation**

Filed by IBM in 2005  
Method and System for Extracting and Visualizing Graph-Structured Relations from Unstructured Text

#### **Knowledge Extraction for e-learning**

Filed by IBM in 2005  
Method and System for Automatically Generating Multilingual Electronic Content from Unstructured Data

**A Syntactic Language Model Based on Incremental CCG Parsing**

In Proceedings of the IEEE 2008 Workshop on Spoken Language Technology, Goa.  
Hany Hassan, Khalil Simaan and Andy Way

**Supertagged Phrase-Based Statistical Machine Translation**

ACL 2007, Prague  
Hany Hassan, Khalil Simaan and Andy Way

**Exploiting Alignment Techniques in MaTrEx: the DCU MT System for IWSLT 2008**

IWSLT 2008, Honolulu, HI  
Yanjun Ma, John Tinsley, Hany Hassan, Jinhua Du and Andy Way

**Language Independent Text Correction using Finite State Automata**

IWSLT 2008, Honolulu, HI  
Proceedings of the 2008 International Joint Conference on Natural Language Processing (IJCNLP, 2008)

**MaTrEx: the DCU Machine Translation System for IWSLT2007**

In Proceedings of IWSLT 2007, Trento, Italy  
Hany Hassan, Yanjun Ma and Andy Way

**Improving Named Entity Translation by Exploiting Comparable and Parallel Corpora**

Ahmed Hassan, Haytham Fahmy and Hany Hassan  
Proceedings of the 2007 Conference on Recent Advances in Natural Language Processing (RANLP, 2007)

**Arabic Cross-Document Person Name Normalization**

Semitic Languages workshop - ACL 2007, Prague  
Walid Magdy, Kareem Darwish, Ossama Emam and Hany Hassan

**BioNoculars: Extracting Protein-Protein Interactions from Biomedical Text**

BioNLP workshop - ACL 2007, Prague  
Amgad Madkour, Kareem Darwish, Hany Hassan, Ahmed Hassan, Ossama Emam

**Syntactic Phrase-Based Statistical Machine Translation**

In Proceedings of the IEEE 2006 Workshop on Spoken Language Translation, Palm Beach, Aruba  
Hany Hassan, Mary Hearne, Andy Way and Khalil Simaan

**Unsupervised Information Extraction Approach Using Graph Mutual Reinforcement**

Proceedings of Empirical Methods for NLP EMNLP, 2006  
Hany Hassan, Ahmed Hassan and Ossama Emam

**Graph Based Semi-Supervised Approach for Information Extraction**

Proceedings of HLT/NAACL 2006  
Hany Hassan, Ahmed Hassan and Sara Noeman

**An Integrated Approach for Arabic-English Named Entity Translation**

ACL 2005, Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages  
Hany Hassan and Jeffrey Sorensen:

**Examining the Effect of Improved Context Sensitive Morphology on Arabic IR**

ACL 2005, Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages  
Kareem Darwish, Hany Hassan and Ossama Emam

**A Statistical Model for Multilingual Entity Detection and Tracking**

Proceedings of HLT-NAACL 2004  
Radu Florian, Hany Hassan, Abraham Ittycheriah, Hongyan Jing, Xiaoqiang Luo, Nicolas Nicolov and Salim Roukos

**Language Model Based Arabic Word Segmentation**

Proceedings of ACL 2003  
Young-Suk Lee, Kishore Papineni, Salim Roukos, Hany Hassan

**TIPS: A Translingual Information Processing System**

Proceedings of HLT-NAACL 2003

Yaser Al-Onaizan, Radu Florian, Martin Franz, Hany Hassan, Young-Suk Lee, J. Scott McCarley, Kishore Papineni, Salim Roukos, Jeffrey Sorensen, Christoph Tillmann, Todd Ward, Fei Xia,

**A Maximum Entropy Approach for Arabic Text Categorization**

MSc thesis, Faculty of Engineering Cairo University, 2001

Hany Hassan

**Awards and Honors**

IBM Software Group Top Talent for Leadership Selected among IBM Software Group top talents candidates for leadership positions.	2008
IBM First Plateau Invention Achievement Award In appreciation and recognition of creative contribution to IBM progress.	2007
IBM Watson Research Center Bravo Award For significant contribution in Information Extraction Research.	2006
IBM Invention Achievement Award For First Patent Application.	2005
Science Foundation of Ireland (SFI) PhD scholarship at Dublin City University.	2005
IBM Technical Achievement Award For outstanding contribution in EMMS development	2001

**Technical Skills**

- Strong combination of applied and academic research skills
- Strong Experience in Statistical Modeling and Machine Learning techniques.
- Highly professional experience in Statistical Machine translation.
- Highly professional experience in Information Extraction and Knowledge discovery
- Highly professional experience in C++ and Object Oriented programming.
- Strong experience in Java, Perl, Shell Scripting and TCL/TK programming.
- Experience in Linux, Unix and Windows systems

**Leadership Skills**

- Demonstrated technical leadership capability by leading engineers to achieve complex projects.
- Managed and coordinated various complex projects.
- Provided guidance and mentorship for junior team members.
- Knowledge of Software Project Management.

**References**

Dr. Salim Roukos  
IBM Watson Research Center, NY, Senior Manager of Multilingual NLP Technologies and IBM CTO of Translation Technologies.

Dr. Ahmed Tantawy  
IBM, CTO - SWG Mega Deals

Prof. Andy Way  
Dublin City University, Leader for Integrated Language Technologies in CNGL.

Dr. Khalil Simaan  
University of Amsterdam, Senior lecturer at the Faculty of Exact Sciences.

Dr. Srinivas Bangalore  
AT&T Research Labs, NJ, Principal Technical Staff Member.

**Contact  
Information**

Work Address:

IBM - Egypt  
Cairo Technology Development Center  
Pyramids Heights Office Park  
Cairo-Alexandria Desert Road, km. 22  
Giza - Egypt

Work Phone:

(+20) 235361447

Mobile Phone

(+20) 10 6066076

E-mail

[hany.hassan2@mail.dcu.ie](mailto:hany.hassan2@mail.dcu.ie)

Webpage:

<http://www.computing.dcu.ie/~hhasan/>