KE5205 TEXT MINING 2018

MTECH KNOWLEDGE ENGINEERING

Fan Zhen Zhen

Cai Yuhao

Institute of Systems Science

National University of Singapore

FAN Zhen Zhen

Senior Lecturer & Consultant, Analytics & Intelligent Systems Practice

issfz@nus.edu.sg

Profile

Zhenzhen has been with Institute of Systems Science, NUS, since 2006. She currently lectures in the Master of Technology programme in the areas of case-based reasoning, text mining, KBS development, hybrid KBS, and formal specification. Prior to joining ISS, she was a senior research engineer at the Institute for Infocomm Research working in the areas of machine translation and natural language processing. Her current research interests lie in text mining and computational linguistics.

Educational Qualifications/Professional Certifications +

Academic and Professional Experience +

Membership and Professional Activities +

Research Interests −

- Text Mining & Data Mining
- Natural Language Processing
- Computational Linguistics
- Machine Translation
- Knowledge Representation

Cai Yuhao

Associate Lecturer & Consultant, Analytics & Intelligent Systems Practice

isscy@nus.edu.sg

Profile

Yuhao is currently a researcher and a consultant in NUS, Institute of Systems Science. His research interest is in statistical data analysis and data visualization. He is also a level 2 Candidate in the CFA Program.

What I Teach

ANALYTICS AND INTELLIGENT SYSTEMS

NICF- Statistics for Business

ANALYTICS AND INTELLIGENT SYSTEMS

NICF- Statistics Bootcamp

ANALYTICS AND INTELLIGENT SYSTEMS

NICF- Text Analytics

STACKUP - STARTUP TECH TALENT DEVELOPMENT

Machine Learning driven Data Science

Adjunct Lecturer

Rajaraman KANAGASABAI

Principal Investigator
Data Mining Department
Institute for Infocomm Research
1 Fusionopolis Way
#21-01 Connexis (South Tower)
Singapore 138632
Email: kanagasa@i2r.a-star.edu.sg
Web: http://datam.i2r.a-star.edu.sg/~kanagasa/

Highlights of best achievements

- Strong research record in Data Analytics with over 10 years of research experience and publications in top ranked and peer-reviewed international journals and conferences.
- Well-experienced in leading teams into premier applied research groups with core IP creation (publications & patents), research collaborations, international recognition, and spinoffs.
- Co-inventor of the Knowle technology that has been spunoff in 2010 as Knorex (www.knorex.com). Core research team member behind the multiple-award winning iAgent - the first multilingual search engine, WebWatch - the key technology behind the successful startup BuzzCity (www.buzzcity.com). Leader of the project that won the **Tan Kah Kee Young Inventor's award 2006** (Merit – Open category).

Education

Ph.D., Indian Institute of Science, Bangalore, India. 1997.
M.Engg., Indian Institute of Science, Bangalore, India. 1989-91. Gold Medalist.
B.Engg., Anna University, Madras, India. 1985-89. First class with Distinction.

Current Research Interests

Data Analytics: Ontology-centric Analytics, Data Integration, Text Analytics, Web X.0 mining
Semantic Technologies: Semantic Data Integration, Semantic Search, Linked Data, Ontology Learning & Population, Semantic Querying
Cloud/Services Computing: Service discovery& composition, Semantic services, Cloud Analytics

Professional Activities / Special Achievements

Teaching:

- Invited Guest Speaker on 'Semantic Technologies' at PREMIA Data Analytics Course
- Invited Guest Lecturer on 'Biological Data Mining' at NUS' Bioinformatics course.

Grants/Funding

- Co-PI, "Knowle G2", Commercialization of Technology Project, (Exploit Technologies funded).
- PI, Ontology-based Contextual Search and Query Engine (Industry funded Research Collaboration Project).
- Intramural A*STAR (I^2R Funding)

Awards

- Leader of the KnowleTracker technology that won the RunnerUp in **'Web Applications Competition'** @iWAS 2009.
- Leader of the project that won the **Tan Kah Kee Young Inventor's award 2006** (Merit – Open category).
- Gold medalist for top academic achievement in Masters course at Indian Institute of Science in 1991.

Course Schedule (Saturdays)

Saturday	Saturday	Saturday	Saturday	Saturday
1 Sept	8 Sept	15 Sept	22 Sept	29 Sept
9:00 – 12:30	9:00 – 12:30	9:00 – 12:30	9:00 – 12:30	9:00 – 12:30
Introduction to Text Mining	Preparing Textual Data for Analysis	Text Categorization	Information Extraction	Link and Trend Analysis
The Text Mining Process	Workshop 2 Text Preparation	Workshop 4	Linguistic and Knowledge Resources	Workshop 7
Lunch	**Lunch**	**Lunch**	**Lunch**	**Lunch**
1:30 – 5:00	1:30 – 5:00	1:30 – 5:00	1:30 – 5:00	1:30 – 5:00
Tools and Solutions for Text Mining	Clustering & Topic Modeling	Workshop 5	Workshop 6	Advanced Topics in Text Mining
Workshop 1 Data Acquisition	Workshop 3			Case study & Workshop 8

Course Schedule (Tuesdays & Thursdays)

Tuesday	Tuesday	Tuesday	Tuesday	Tuesday
4 Sept	11 Sept	18 Sept	25 Sept	2 Oct
18:30 – 22:00	18:30 – 22:00	18:30 – 22:00	18:30 – 22:00	18:30 – 22:00
Introduction to Text Mining	Preparing Textual Data for Analysis	Text Categorization	Information Extraction	Link and Trend Analysis
The Text Mining Process	Workshop 2 Text Preparation	Workshop 4	Linguistic and Knowledge Resources	Workshop 7
Thursday	**Thursday**	**Thursday**	**Thursday**	**Thursday**
6 Sept	13 Sept	20 Sept	27 Sept	4 Oct
18:30 – 22:00	18:30 – 22:00	18:30 – 22:00	18:30 – 22:00	18:30 – 22:00
Tools and Solutions for Text Mining	Clustering & Topic Modeling	Workshop 5	Workshop 6	Advanced Topics in Text Mining
Workshop 1 Data Acquisition	Workshop 3			Case study & Workshop 8

Reflection

At the end of this course, I should be able to…

1. _____

_____

_____

2. _____

_____

3. _____

Objectives of this course

At the end of this course, you can:

- Describe and follow the general text mining process to discover insights such as relations, patterns, trends, etc., from textual data
- Understand the concepts and be able to apply the techniques of major text mining tasks such as concept extraction, text categorization, clustering, topic modelling, association analysis, etc.
- Understand the importance of domain-specific lexical and knowledge resources such as vocabularies, thesaurus, taxonomies, ontology, rules, etc., and be able to develop such resources to improve the mining results

KE-TM Course Assessment

- Continual Assessment (50 marks)
  - In-class case study: **15** marks
  - CA Project (team based): **35** marks
- Final Exam (open book) **50** marks