

DATA MINING GROUP

ADVISOR: PEERAPON VATEEKUL, PH.D.

DEPT. OF COMPUTER ENGINEERING

FACULTY OF ENGINEERING, CHULALONGKORN UNIVERSITY

Members

Advisor

- Peerapon Vateekul, Ph.D.
 - Email: peerapon.v@chula.ac.th

Graduate Members

- **Natchanon Phachongkitphiphat (Nat)**
Master of Engineering: Computer Engineering (CM)
Expected Graduation: 2015
- **Suthipong Daengduang (Top)**
Master of Engineering: Computer Science (CS)
- **Teerawit Choeikiwong (Beam)**
Master of Science: Software Engineering (SE)
Expected Graduation: 2015
- **Tanawat Limungkura (Mum)**
Master of Engineering: Computer Science (CS)

Undergraduate Members

- B.Eng (CP) Expected Graduation: 2015
 - Nuttapon Pattanavijit (Nut)
 - Vibhavee Trairattanapa (Vee)
 - Kankawin Kowsrihawat (K)
 - Panida Nimnual (Chompoo)
 - Piyawat Lertvittayakumjorn (Tui)

Projects

Large-scaled Hierarchical Multi-label Classification

- **Objective:** Propose new methods for addressing hierarchical multi-label classification problem which outperform the existing standard algorithm in terms of time and accuracy.
- **Challenges:** Large datasets, Multi-label, Tree/DAG class structure, Full/partial depth labeling



Dataset	#Train docs	#Test docs	#Leaf classes	#Features	Max Depth
Wiki Medium	456,886	81,262	36,504	346,299	12
Wiki Large	2,365,436	452,167	325,056	1,617,899	729

Proposed method:



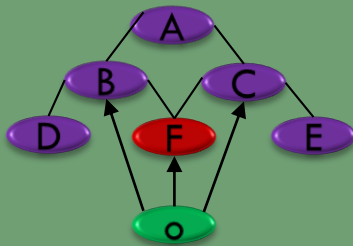
Hierarchical Multi-Label Classification for Complex Hierarchical Structure

Dataset: 
Medical Document Database
12,628,968 Document

Class Hierarchy:
Medical Subject
Heading 26,831 Words

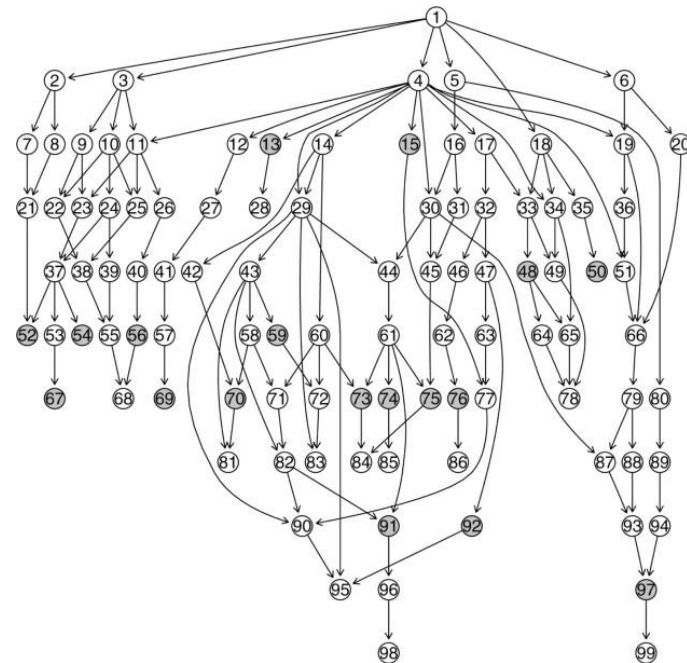


Hierarchical Multi-Label



o belong to f then o belong to B,C

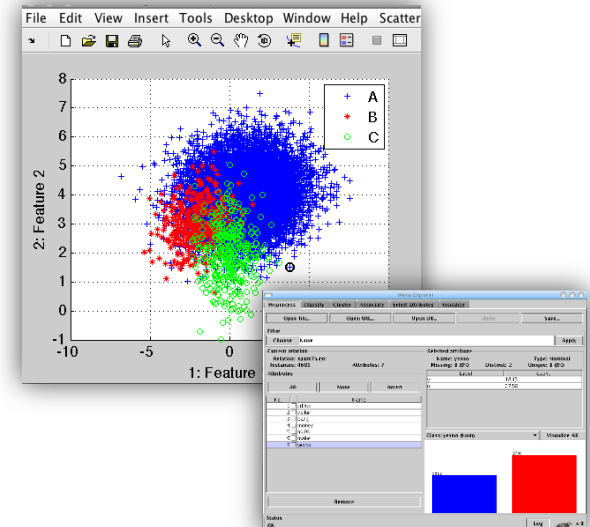
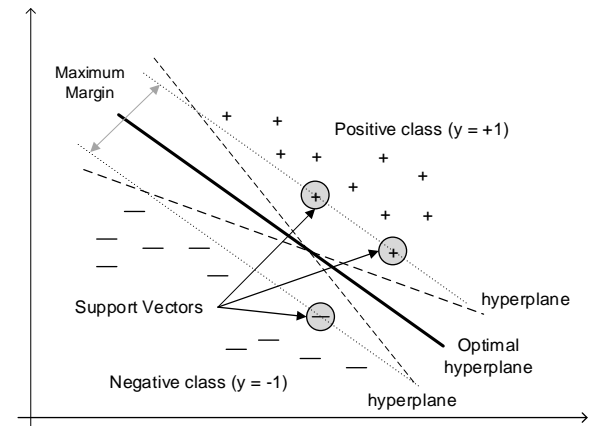
- **Challenges:** Complex Hierarchical Structure, Directed Acyclic Graph(DAG) and Non-Mandatory Leaf Node(NMLN)



Software Defect Prediction Using Unbiased Support Vector Machine



- **What ?** This research aims to achieve high defect prediction performance in terms of PD, PF, F1, and G-mean.
- **Why?** Prediction performance was limited due to the imbalanced issue since the number of defected modules is very small comparing to that of non-defected modules.
- **How?** By employing unbiased Support Vector Machine called “R-SVM” to tackle the imbalanced issue. In R-SVM, the separation hyperplane is adjusted to reduce a bias from the majority class (non-defect).



Gist Extraction System on Thai Supreme Court Cases

- ❑ **Goal:** Prototype Information Extraction and Automatic Summarization System for Thai Supreme Court Judgment
- ❑ **Benefit:** Help on matters of law summarization and useful judgment selection for legal research
- ❑ Create as Web-based Application



JUDGEDOLL

ระบบสืบค้นใจความสำคัญจากคำพิพากษาศาลฎีกา

Download File >>>

.docx
 .doc
 .pdf
 .txt

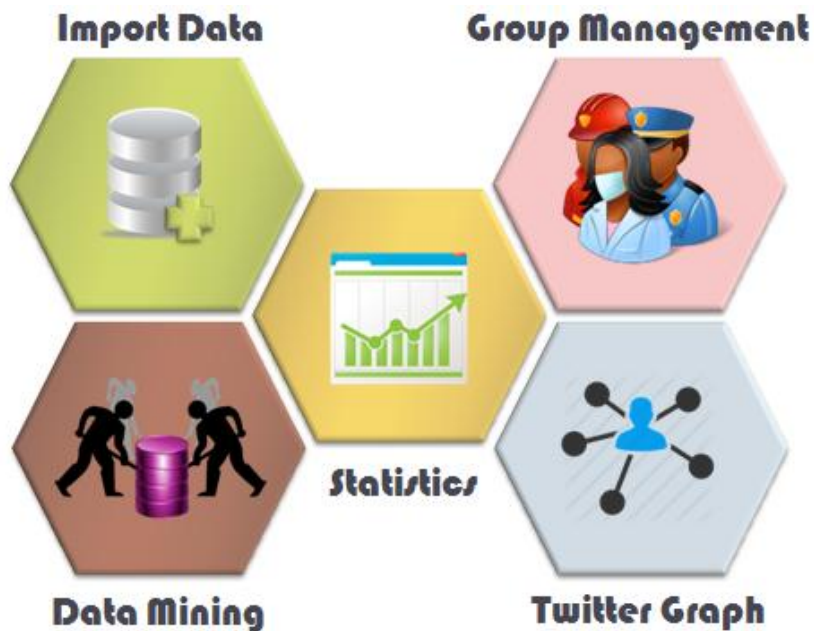
ประเภทคดี	แพ่ง
ตุลาการในคดี	<div style="display: flex; justify-content: space-between;"> พ่วงคู่ส่วนจำกัด สิทธิสิทธิ์ถือครองบางแก้ว (รับลา) โจทก์ </div> นายเฉลิมพล ศิริเทพย์หรือบุตรในพระชน ที่ ๑ นายบ๊วย ศิริพนมพลหรือบุตรในพระชน ที่ ๒
เรื่อง	ลาภมิควรได้ (รับมรดกคดี)
ผู้ยื่นฎีกา	จำเลยทั้งสอง

ย้อนกลับ

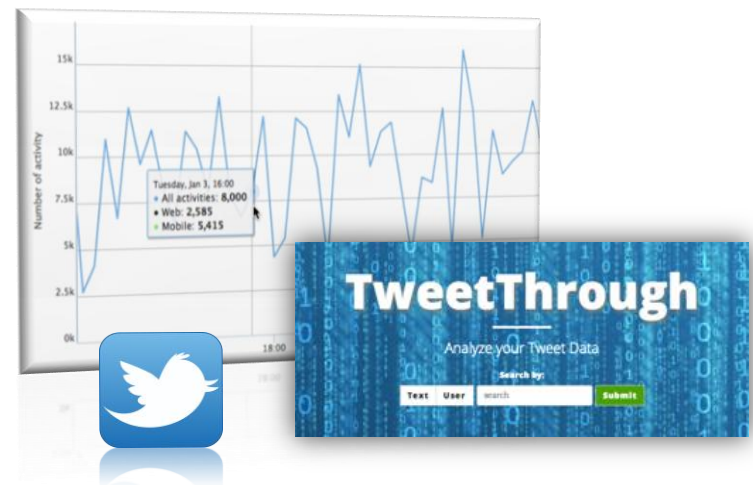
การที่โจทก์และจำเลยทั้งสองตกลงกันทำสัญญาประนีประนอมยอมความคดีศาลอุทธรณ์ และศาลได้พิพากษายกฟ้องไปแล้วนั้น โจทก์และจำเลยทั้งสองยังมีเจตนาต่อกันที่จะให้สัญญาประนีประนอมยอมความสำหรับเรื่อง ประสงค์ของทั้งสองฝ่าย ข้อความสำคัญของสัญญาประนีประนอมยอมความดังกล่าวในข้อ ๒ คือ จำเลยทั้งสองจะชำระเงินจำนวน ๓,๓๐๐,๐๐๐ บาท แก่โจทก์พร้อมดอกเบี้ยร้อยละ ๘.๕ ต่อปี นับแต่วันที่ ๑๐ สิงหาคม ๒๕๕๑ เป็นต้นไป การ ชำระหนี้สำหรับ ค่าจำเลยทั้งสองมีผู้ชำระหนี้ดังกล่าวแก่โจทก์ ผิดและเป็นเหตุให้เกิดคดีฟ้องคดีตามสัญญาข้อ ๒ นี้ ส่วนที่มีข้อความว่า โจทก์จะขอฟ้องโจทก์ชำระเงินดังกล่าวแก่โจทก์ เมื่อจำเลยทั้งสองสามารถชดเชยได้ในคดีข้อที่ ๒๖๕๔ คำพิพากษา จำเลยทั้งสองไม่ชำระคืน สัญญาประนีประนอมยอมความ ศาลอุทธรณ์จึงพิพากษาว่าโจทก์ฟ้องโจทก์และจำเลยทั้งสองมีสัญญาประนีประนอมยอมความผิด โจทก์สามารถ ชำระคืนเงินในคดีชำระหนี้ และเป็นสัญญาที่มิได้ทำขึ้นประนอมยอมความการชำระหนี้ไว้ จำเลยทั้งสองต้องชำระหนี้ตามสัญญาประนีประนอมยอมความภายในระยะเวลาอันสมควร การที่จำเลยทั้งสองไม่ชำระหนี้ตามสัญญาประนีประนอม

Twitter Data Dispersion Analysis

- **Objective:** Apply suitable data mining techniques for analyzing dispersion characteristics of Twitter data. Also, develop web application processing and reporting results for user.



- **Focus on:** Political Domain – Influence User, Fake news



Data Quality Management for Hydrological Data

- **Objective:** Devise a system to capture and notify inconsistency in hydrological data. Moreover, research new methods to enhance detecting flawed cumulative rain level data.
- **Common Tasks:** missing value imputation, outlier detection, data consistency check, and homogenization

