

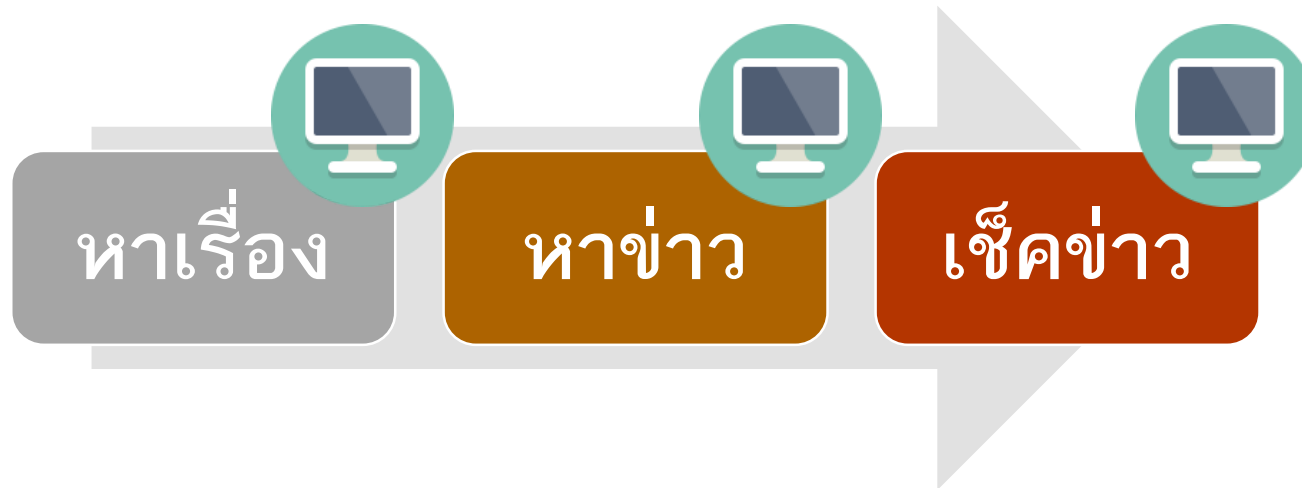
Future Possibilities of Rumor Detection System



ปิยวัฒน์ เลิศวิทยากำจร

ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย

ความเป็นไปได้ในอนาคต



เทคนิคหาเรื่อง

- Twitter REST API: GET trends/place

```
{
  "events": null,
  "name": "#10PersonasQueExtra\u00f1oMucho",
  "promoted_content": null,
  "query": "%2310PersonasQueExtra%C3%B1oMucho",
  "url": "http://twitter.com/search/?q=%2310PersonasQueExtra%C3%B1oMucho"
},
{
  "events": null,
  "name": "Apple $1.5",
  "promoted_content": null,
  "query": "%22Apple%20$1.5%22",
  "url": "http://twitter.com/search/?q=%22Apple%20$1.5%22"
},
{
  "events": null,
  "name": "Zelko",
  "promoted_content": null,
  "query": "Zelko",
  "url": "http://twitter.com/search/?q=Zelko"
},
}
```

ข้อดี

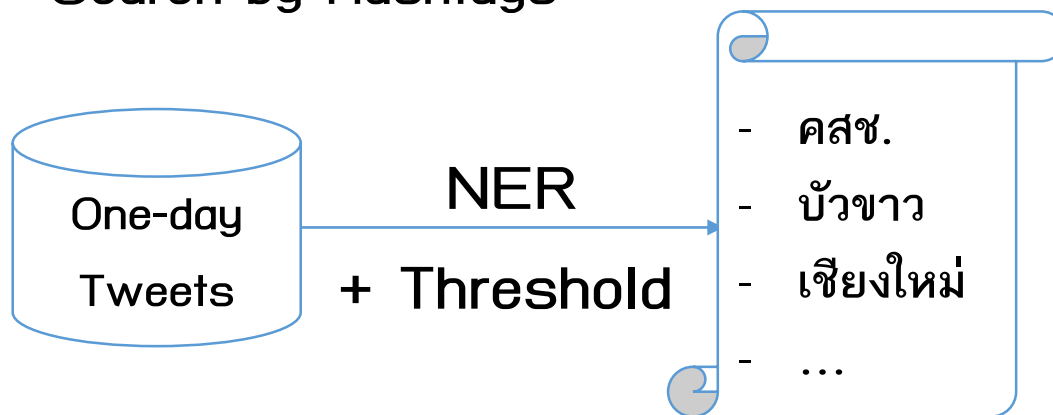
- ด่วน ทันเหตุการณ์
- สะดวก
- Group Trend ที่คล้ายกันอยู่ด้วยกัน

ข้อเสีย

- ถ้าจำนวนทวีตไม่เยอะพออาจตรวจจับไม่ได้
- ไม่รู้วิธีการเลือก trend ของ Twitter

เทคนิคหาเรื่อง

- Named Entity Recognition (NER)
- Search by Hashtags



ข้อดี

- หัวข้อไม่หล่นหาย
- กำหนดความถี่ได้เอง

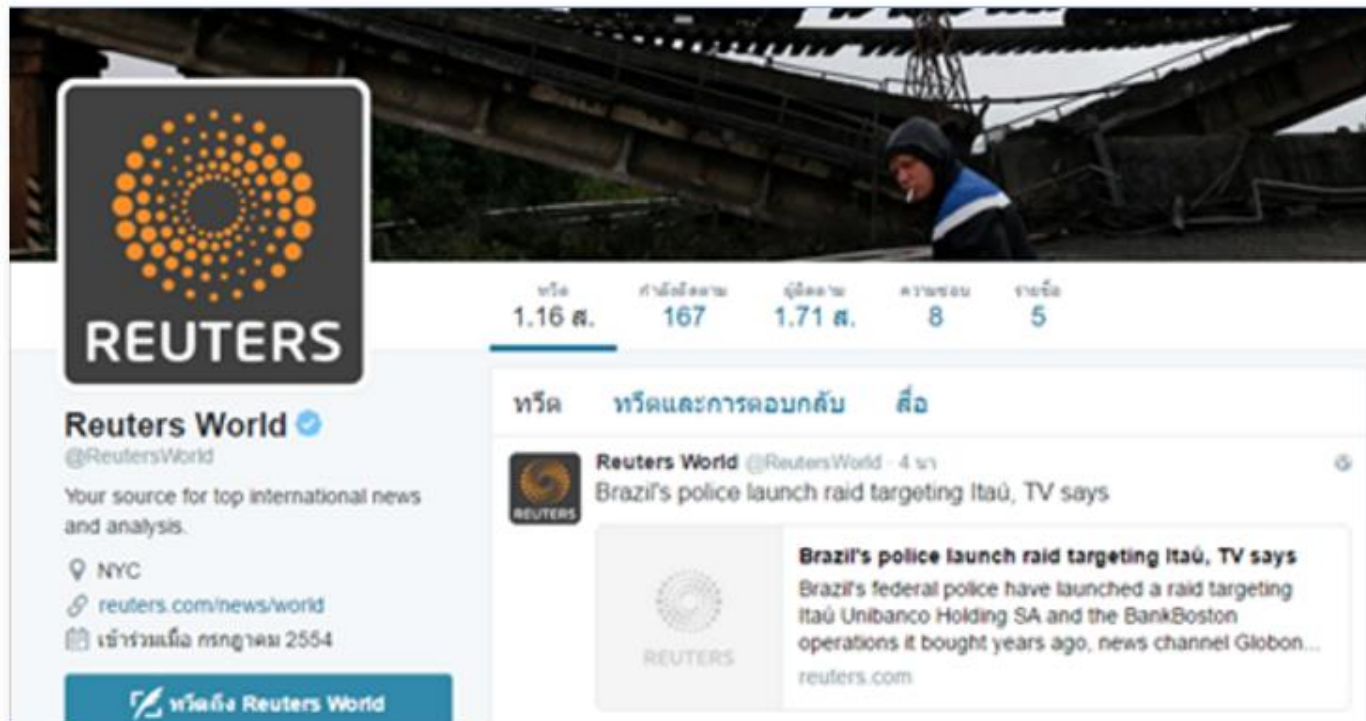
ข้อเสีย

- อาจมีหัวข้อซ้ำซ้อน
- ไม่มีซอฟต์แวร์อัตโนมัติ
- อาจต้องใช้เกณฑ์อื่นช่วย เช่น Retweet Rate

รายละเอียดวิทยานิพนธ์

ชื่อวิทยานิพนธ์	การรู้จำและการจำแนกประเภทของชื่อเฉพาะภาษาไทย NAMED ENTITY RECOGNITION AND CLASSIFICATION IN THAI.
ชื่อนิสิต	สุฤดี ฉัตรไตรมงคล Surudee Chattrimongkol
ชื่ออาจารย์ที่ปรึกษา	ผศ.ดร. วิโรจน์ อรุณมานะกุล Asst.Prof. Wirote Aroonmanakun, Ph.D.
ชื่อสถาบัน	จุฬาลงกรณ์มหาวิทยาลัย. บัณฑิตวิทยาลัย Chulalongkorn University. Bangkok (Thailand). Graduate School.
ระดับปริญญาและรายละเอียดสาขาวิชา	วิทยานิพนธ์มหาบัณฑิต. อักษรศาสตร์ (ภาษาศาสตร์) Master. Arts (Linguistics)
ปีที่จบการศึกษา	2548

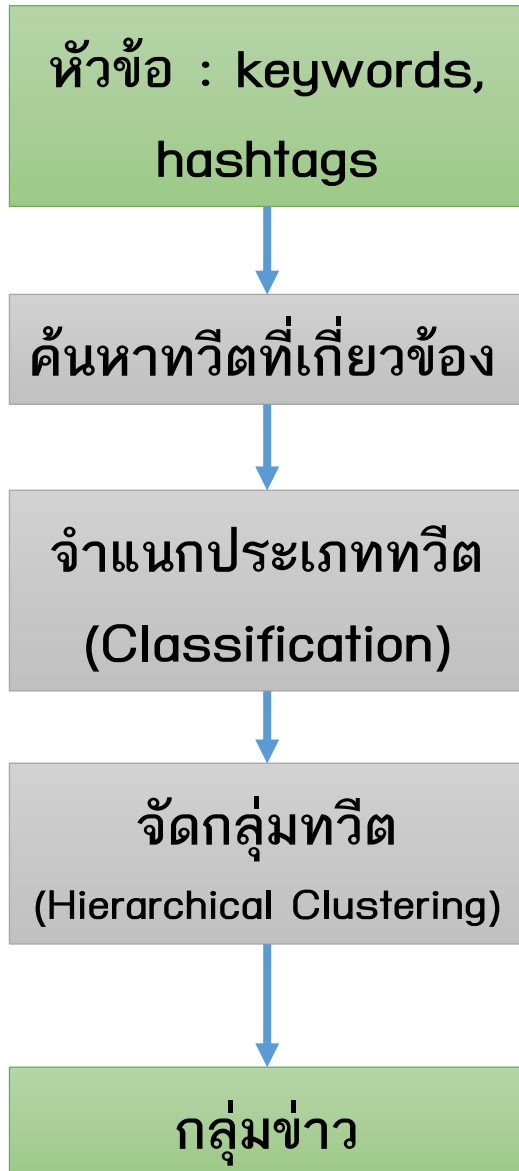
เทคนิคหาเรื่อง



Reuters สร้างอัลกอริทึมของตัวเองเพื่อดูว่ามีอะไรใน Twitter มาเป็นข่าวด่วนได้บ้าง | Blognone

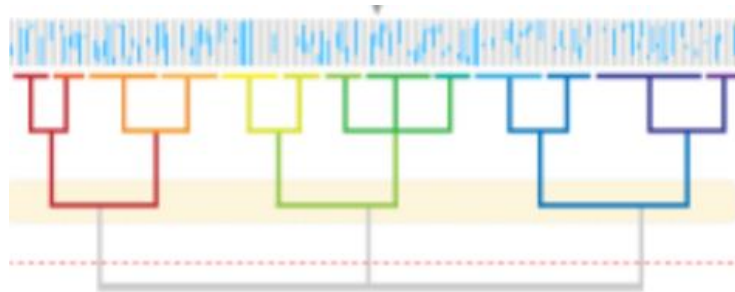
WWW.BLOGNONE.COM

ขั้นตอนหาข่าว



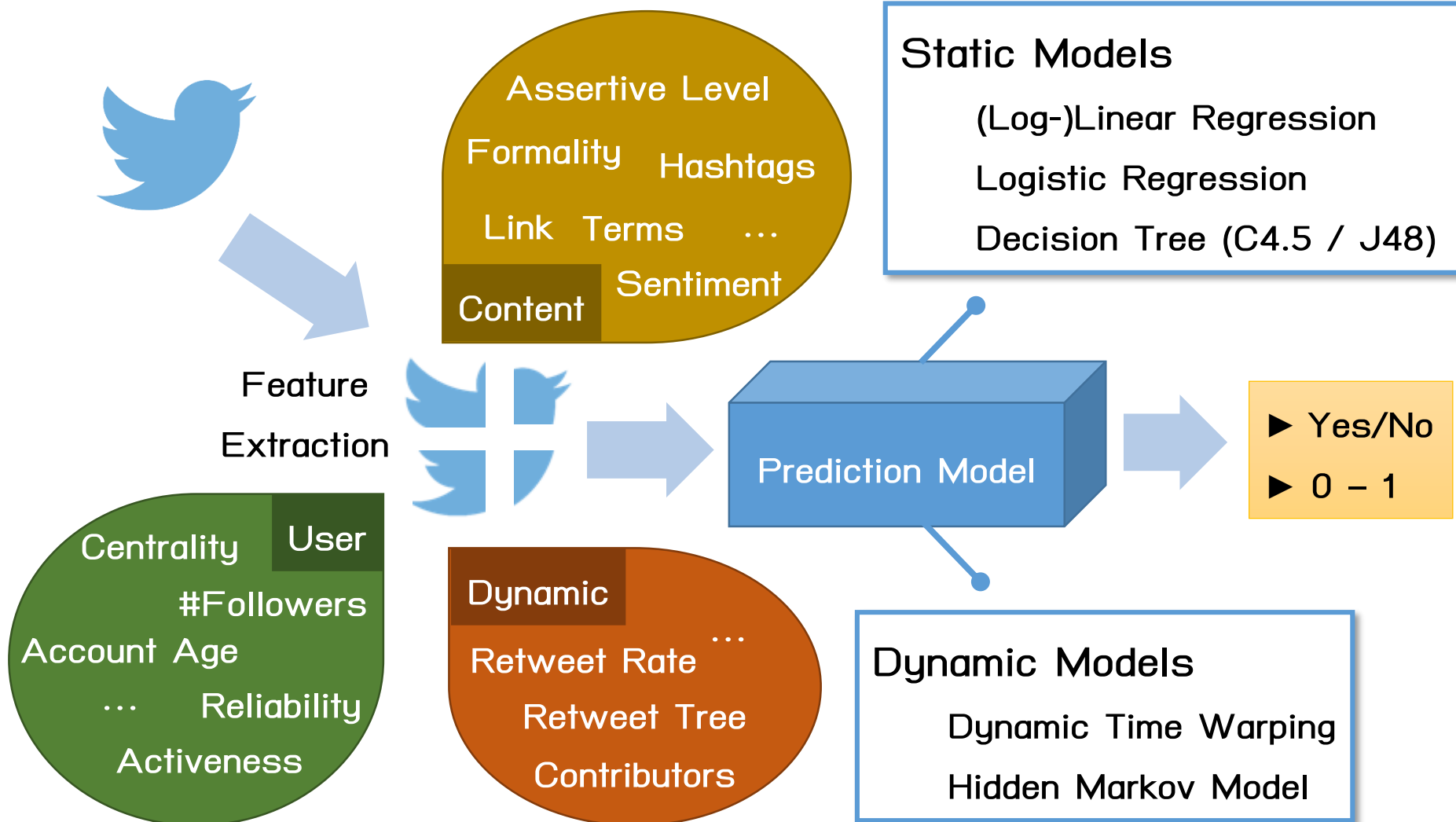
- Twitter REST API
- In Database

Assertion, Recommendation, Expression, Question, Request, Miscellaneous



Assertion Group1
Assertion Group2
Assertion Group3

ใช้ข่าว >> Classification



เช็คว่า >> Classification

Features

- เนื้อหาและการใช้ภาษา
 - คำที่ใช้, ความเป็นทางการ, น้ำหนัก, ทัศนคติ, ลี้นัก, hashtags
- ผู้ทวิตและเครือข่ายการติดตาม
 - อายุการใช้งาน, จำนวนผู้ติดตาม, ความน่าเชื่อถือ, การโต้ตอบ
- การส่งต่อทวิต
 - จำนวนและอัตราการรีทวิต, ลักษณะของ Retweet Tree
- และอื่น ๆ

Methods

- Static
 - (Log-)Linear Regression
 - Logistic Regression
 - Decision Tree (C4.5 / J48)
- Dynamic
 - Dynamic Time Warping
 - Hidden Markov Model



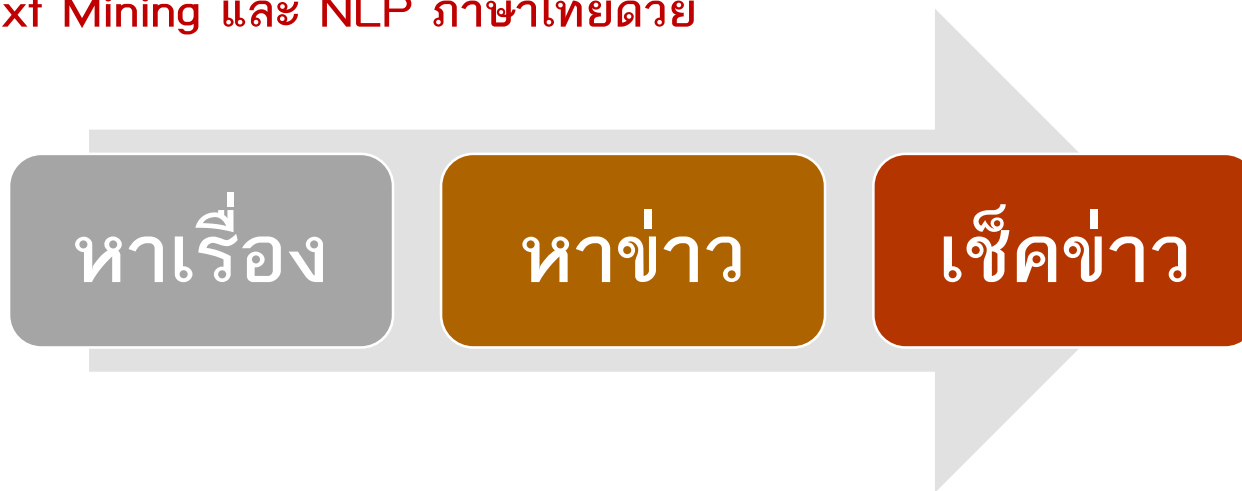
Dataset

ความเป็นไปได้ในอนาคต

- เราจะหาข่าวที่ควรตรวจสอบได้อย่างอัตโนมัติหรือไม่ ? **เป็นไปได้**
- เราจะจำแนกทวีตออกเป็นข่าวจริงกับข่าวลืออัตโนมัติได้หรือไม่ ?

เป็นไปได้ยากกว่า เนื่องจาก Training data มีน้อย และการตัดสินใจข่าวลือทำได้ยาก
แม้แต่มนุษย์ทำเองก็ตาม

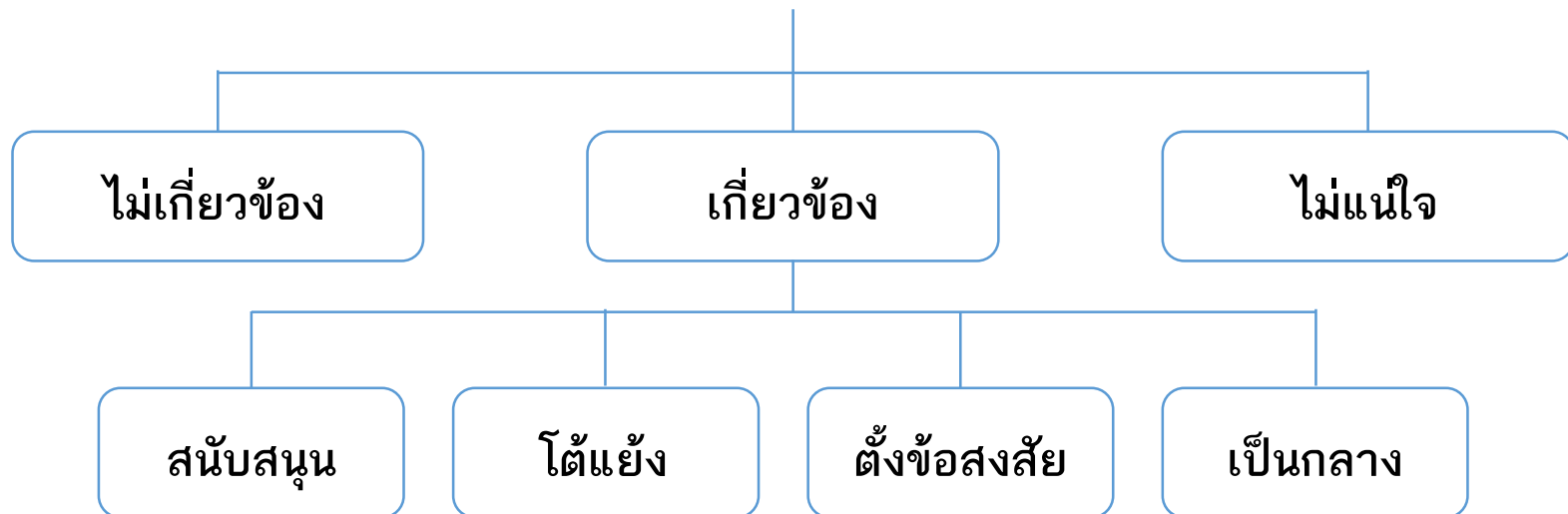
- ทำบทวิเคราะห์**ภาษาไทย**ได้หรือไม่ ? **ทำได้** แต่ประสิทธิภาพขึ้นกับความก้าวหน้า
ด้าน Text Mining และ NLP ภาษาไทยด้วย



ชุดข้อมูล (Datasets)

- (V.Qazvinian, 2011)

Name	Rumor	Regular Expression Query	#tweets
obama	Is Barack Obama muslim?	Obama & (muslim islam)	4975
airfrance	Air France mid-air crash photos?	(air.france air france) & (photo pic pix)	505
cellphone	Cell phone numbers going public?	(cell cellphone cell phone)	215
michelle	Michelle Obama hired too many staff?	staff & (michelle obama first lady 1st lady)	299
palin	Sarah Palin getting divorced?	palin & divorce	4423



ชุดข้อมูล (Datasets)

- (S.Vosoughi, 2015)

Source/Event	False Rumors	True Rumors	Total
2013 Boston Marathon Bombings	16	6	22
2014 Ebola Pandemic	11	9	20
2014 Ferguson Unrest	10	7	17
Snopes.com & Factcheck.org	76	74	150
• All	113	96	209

other stories of unknown or questionable origin.

- Factcheck.org : documents inaccurate and misleading claims, verifies online rumors

References

- Vosoughi, Soroush. *Automatic detection and verification of rumors on Twitter*. Diss. Massachusetts Institute of Technology, 2015.
- Hamidian, Sardar, and Mona Diab. "Rumor Detection and Classification for Twitter Data."
- Takahashi, Tetsuro, and Nobuyuki Igata. "Rumor detection on twitter." *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on*. IEEE, 2012.
- Qazvinian, Vahed, et al. "Rumor has it: Identifying misinformation in microblogs." *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011.
- <https://www.blognone.com/node/87784>
- <https://dev.twitter.com/rest/reference/get/trends/place>