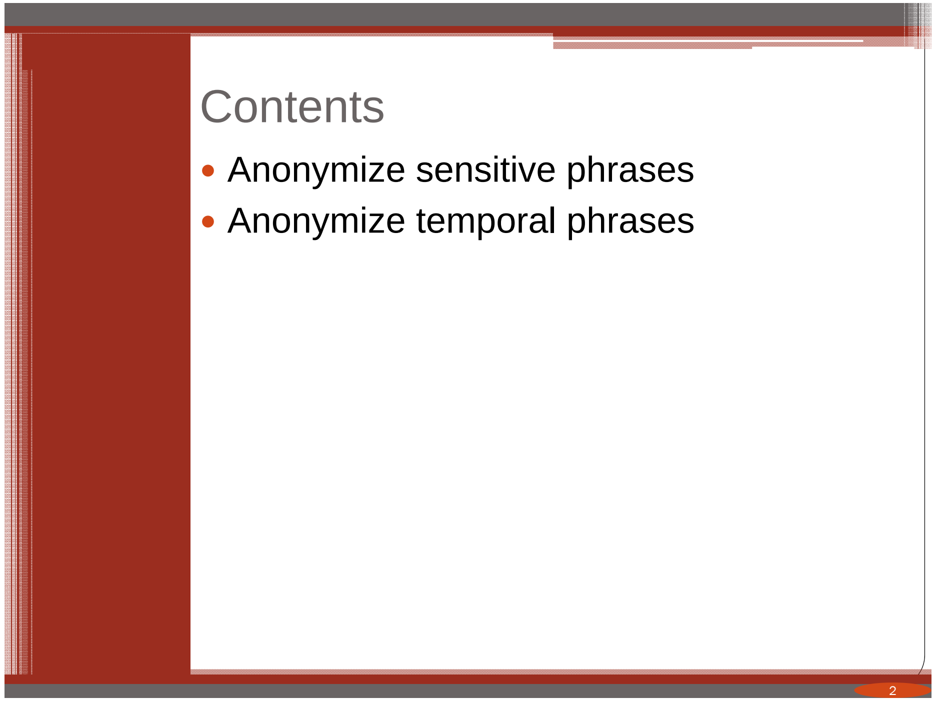
国立大学法人
総合研究大学院大学
The Graduate University for Advanced Studies [SOKENDAI]

大学共同利用機関法人 情報・システム研究機構
NII 国立情報学研究所
National Institute of Informatics

Anonymizing Sensitive Information of Text Posted on Social Networking Services

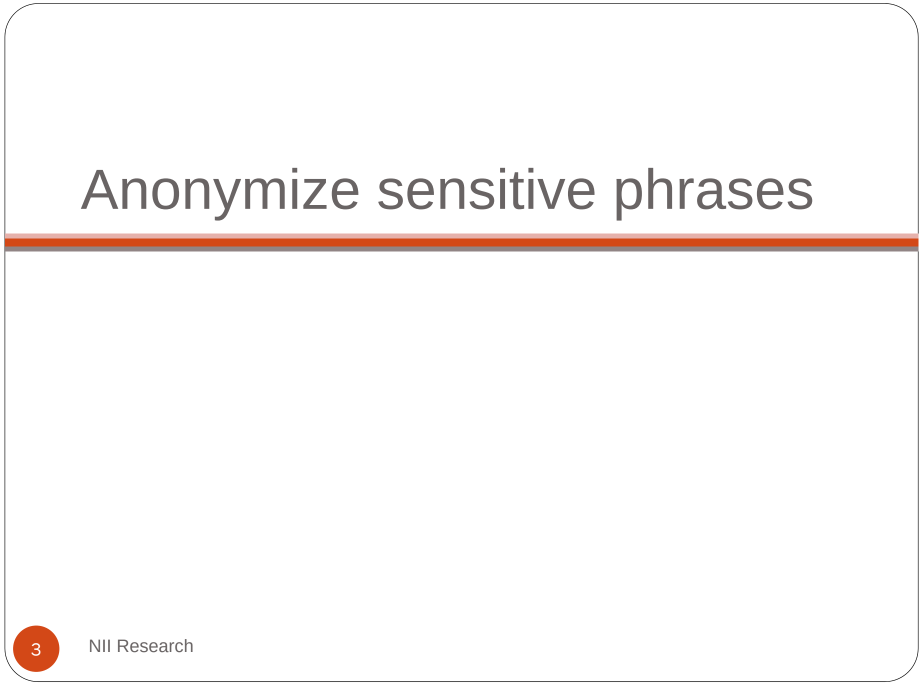
Nguyen Son Hoang Quoc
Isao Echizen

International Workshop on the 6th Information Systems for Social Innovation (ISSI)
February 4, 2014.

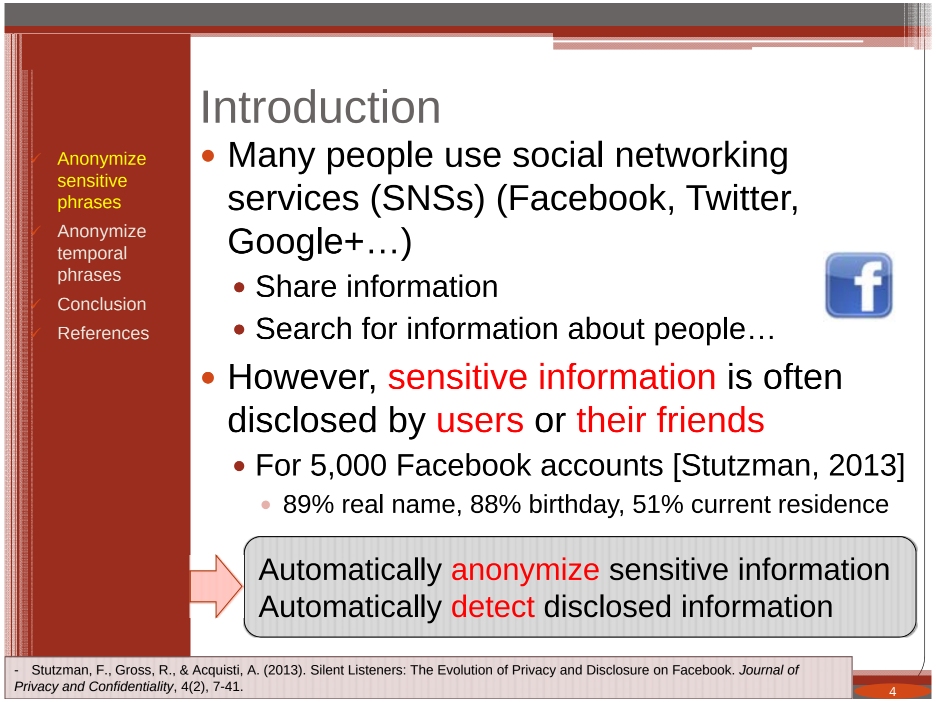


Contents

- Anonymize sensitive phrases
- Anonymize temporal phrases



Anonymize sensitive phrases



- ✓ Anonymize sensitive phrases
- ✓ Anonymize temporal phrases
- ✓ Conclusion
- ✓ References

Introduction

- Many people use social networking services (SNSs) (Facebook, Twitter, Google+…)
  - Share information
  - Search for information about people…
- However, sensitive information is often disclosed by users or their friends
  - For 5,000 Facebook accounts [Stutzman, 2013]
    - 89% real name, 88% birthday, 51% current residence

Automatically anonymize sensitive information
Automatically detect disclosed information

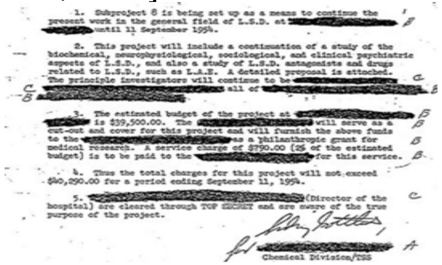
- Stutzman, F., Gross, R., & Acquisti, A. (2013). Silent Listeners: The Evolution of Privacy and Disclosure on Facebook. *Journal of Privacy and Confidentiality*, 4(2), 7-41.

Related Work

Anonymous Text

- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

- **Remove** all sensitive information in texts [Kokkonakis,2007]



1. Subproject 8 is being set up as a means to continue the present work in the general field of L.S.D. at [redacted] B
[redacted] until 11 September 1954.

2. This project will include a continuation of a study of the biochemical, neurophysiological, sociological, and clinical psychiatric aspects of L.S.D., and also a study of L.S.D. antagonists and drugs related to L.S.D., such as L.A.E. A detailed proposal is attached. The principle investigators will continue to be [redacted] C
C [redacted] all of [redacted] B
B [redacted]

3. The estimated budget of the project at [redacted] B
[redacted] is $39,500.00. The [redacted] will serve as a B
cut-out and cover for this project and will furnish the above funds to the [redacted] as a philanthropic grant for B
medical research. A service charge of $790.00 (2% of the estimated budget) is to be paid to the [redacted] for this service. B

4. Thus the total charges for this project will not exceed $40,290.00 for a period ending September 11, 1954.

5. [redacted] (Director of the C
hospital) are cleared through TOP SECRET and are aware of the true purpose of the project.

for [redacted] A
Chemical Division/TSS

➔ **Not natural** after suppression

Anonymize text to be posed on SNS by generalizing sensitive phrases

Detecting Disclosure in Text

- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

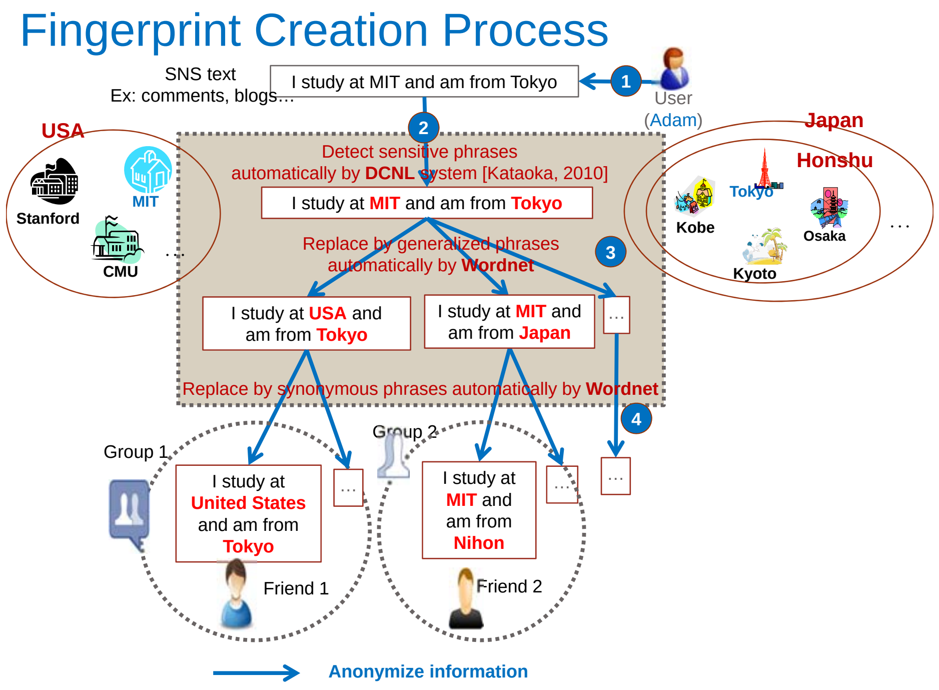
- Use **synonyms** to create a text fingerprint
- Example:
  - Input:
    - You **can insert** a 9 volt battery in the clock radio.
  - Output:
    - F_1: You can **enter** a 9 volt battery in the clock radio.
    - F_2: You **may** insert a 9 volt battery in the clock radio.

➔ **Does not anonymize** the information

Use both synonymization and generalization to anonymize sensitive information to be posted on SNS

Zheng, X., Huang, L., Chen, Z., Yu, Z., Yang, W.: Hiding information by context-based synonym substitution. Proceedings of 8th International Workshop on Digital forensics and Watermarking pp. 162-169 (2009)

Our Algorithm



Fingerprint Creation Process

SNS text
Ex: comments, blogs…

I study at MIT and am from Tokyo

1 User (Adam)

2 Detect sensitive phrases automatically by **DCNL** system [Kataoka, 2010]

I study at **MIT** and am from **Tokyo**

3 Replace by generalized phrases automatically by **Wordnet**

I study at **USA** and am from **Tokyo**

I study at **MIT** and am from **Japan**

…

Replace by synonymous phrases automatically by **Wordnet**

4

USA: Stanford, MIT, CMU, …

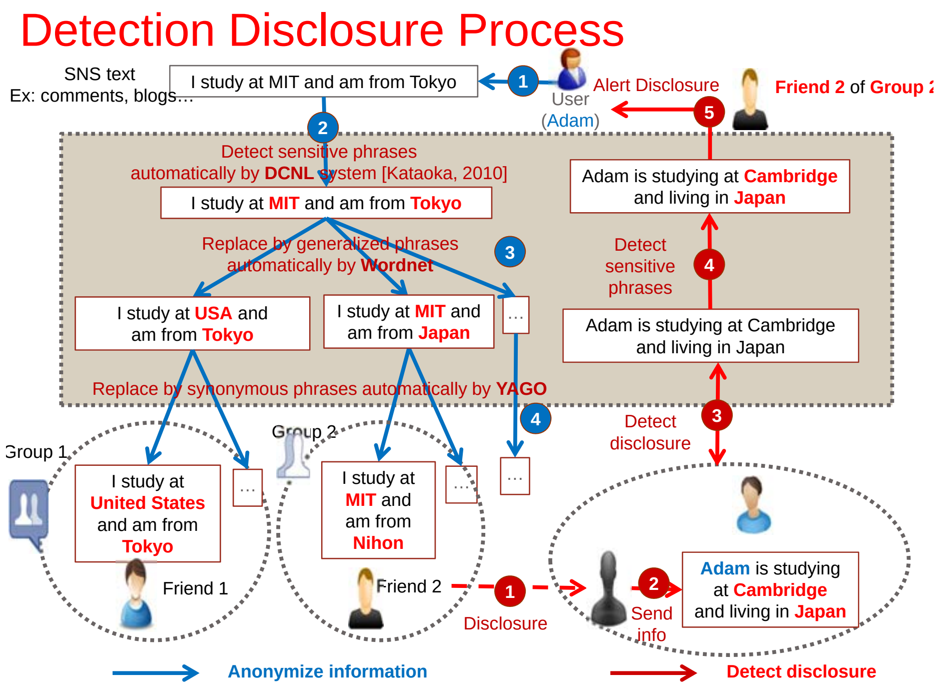
Japan: Honshu: Tokyo, Kobe, Osaka, Kyoto, …

Group 1: I study at **United States** and am from **Tokyo** … Friend 1

Group 2: I study at **MIT** and am from **Nihon** … Friend 2

…

Anonymize information



Detection Disclosure Process

SNS text
Ex: comments, blogs…

I study at MIT and am from Tokyo

1 User (Adam)

2 Detect sensitive phrases automatically by **DCNL** system [Kataoka, 2010]

I study at **MIT** and am from **Tokyo**

3 Replace by generalized phrases automatically by **Wordnet**

I study at **USA** and am from **Tokyo**

I study at **MIT** and am from **Japan**

…

Replace by synonymous phrases automatically by **YAGO**

4

Group 1: I study at **United States** and am from **Tokyo** … Friend 1

Group 2: I study at **MIT** and am from **Nihon** … Friend 2

…

1 Disclosure

2 Send info

Adam is studying at **Cambridge** and living in **Japan**

3 Detect disclosure

Adam is studying at Cambridge and living in Japan

4 Detect sensitive phrases

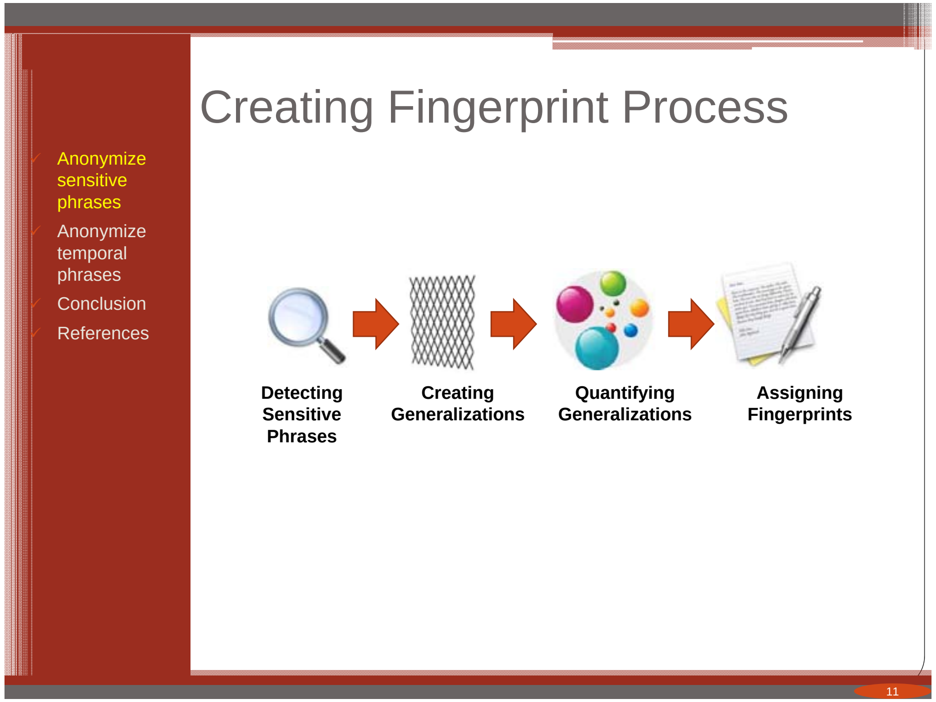
Adam is studying at **Cambridge** and living in **Japan**

5 Alert Disclosure

Friend 2 of **Group 2**

Anonymize information

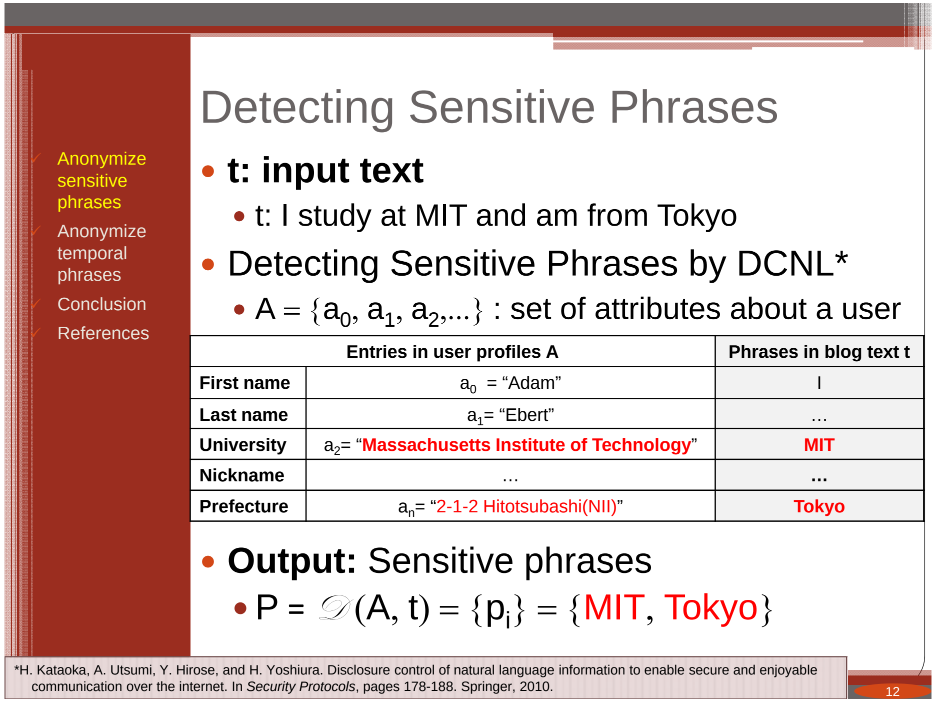
Detect disclosure



Creating Fingerprint Process

- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

Detecting Sensitive Phrases → Creating Generalizations → Quantifying Generalizations → Assigning Fingerprints



Detecting Sensitive Phrases

- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

- **t: input text**
  - t: I study at MIT and am from Tokyo
- Detecting Sensitive Phrases by DCNL*
  - $A = \{a_0, a_1, a_2, \ldots\}$: set of attributes about a user

Entries in user profiles A		Phrases in blog text t
First name	a_0 = “Adam”	I
Last name	a_1= “Ebert”	…
University	a_2= “**Massachusetts Institute of Technology**”	**MIT**
Nickname	…	**…**
Prefecture	a_n= “2-1-2 Hitotsubashi(NII)”	**Tokyo**

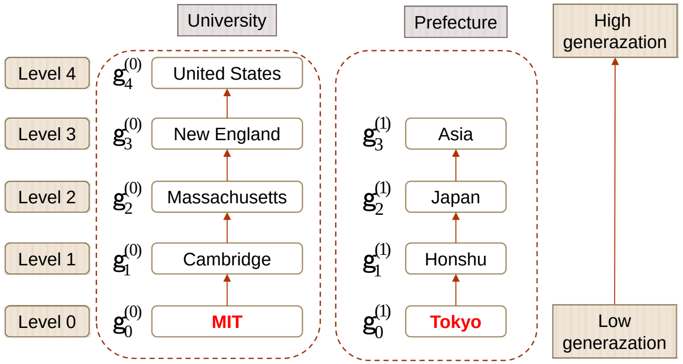
- **Output:** Sensitive phrases
  - $P = \mathscr{D}(A, t) = \{p_i\} = \{$MIT, Tokyo$\}$

*H. Kataoka, A. Utsumi, Y. Hirose, and H. Yoshiura. Disclosure control of natural language information to enable secure and enjoyable communication over the internet. In *Security Protocols*, pages 178-188. Springer, 2010.

Anonymize sensitive phrases
Anonymize temporal phrases
Conclusion
References

Creating Generalization Schemas*

- **Input:** P = {p_i} = {MIT, Tokyo}
- **Output:** Generalization Schemas $G^{(i)}$
  - $G^{(i)} = \mathscr{G}(p_i) = \{\ g_j^{(i)}\ \}$



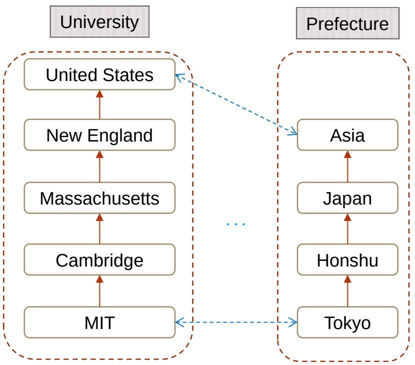
* C. Fellbaum. Wordnet. In *Theory and Applications of Ontology: Computer Applications*, pages 231-243. Springer Netherlands, 2010.

Anonymize sensitive phrases
Anonymize temporal phrases
Conclusion
References

Creating Generalization Schemas*

Input
Generalization Schemas

Output
All possible combined generalizations



Generalizations
{MIT, Tokyo}
{MIT, Honshu}
{MIT, Japan}
{MIT, Asia}
{Cambridge, Tokyo}
{Cambridge, Honshu}
{Cambridge, Japan}
{Cambridge, Asia}
….

* C. Fellbaum. Wordnet. In *Theory and Applications of Ontology: Computer Applications*, pages 231-243. Springer Netherlands, 2010.

Anonymize sensitive phrases
Anonymize temporal phrases
Conclusion
References

Quantifying Generalizations by Modified Discernability Metric DM*

- DM* metric quantifies information loss.*
- The higher the value, the greater the privacy

Low privacy

High privacy

High priority group

Low priority group

Generalization	DM* (sorted)	Group
{MIT, Tokyo}	0.000000E+00	Family
{Cambridge, Tokyo}	1.200930E+09	Best Friends
{MIT, Honshu}	5.967829E+14	Teachers
{Cambridge, Honshu}	5.967841E+14	Students
{MIT, Japan}	**1.117850E+15**	**Friends**
{Cambridge, Japan}	1.117851E+15	Public
…	…	…

Ex: Assign $\{g_0^{(0)}, g_1^{(2)}\}$ **{MIT, Japan}** for **Friends** group

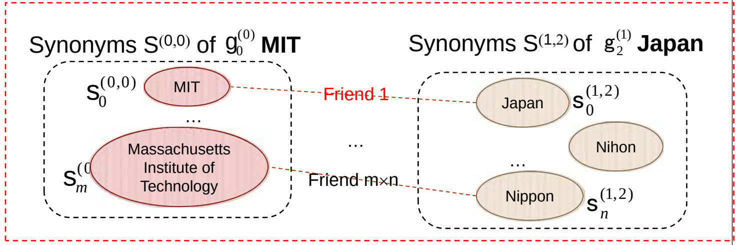
Hoang-Quoc Nguyen-Son, Minh-Triet Tran, Tien-Dung Tran, Hiroshi Yoshiura, Sonehara Noboru, and Isao Echizen, "Automatic Anonymous Fingerprinting of Text Posted on Social Networking Services", *Proc. of the 11th International Workshop on Digital-Forensics and Watermarking (IWDW 2012)*, LNCS, pp. 410-424, Springer (October 2012)

Anonymize sensitive phrases
Anonymize temporal phrases
Conclusion
References

Assignment by Friends

- $S^{(i,j)} = \mathscr{S}(\ g_j^{(i)}\ ) = \{\ s_k^{(i,j)}\ \}$: set of synonyms

Ex: Assign $\{g_0^{(0)}, g_1^{(2)}\}$ **{MIT, Japan}** for **Friends** group


Synonyms $S^{(0,0)}$ of $g_0^{(0)}$ MIT
Synonyms $S^{(1,2)}$ of $g_2^{(1)}$ Japan
$s_0^{(0,0)}$
MIT
…
$s_m^{(0}$
Massachusetts Institute of Technology
Friend 1
…
Friend m×n
Japan
$s_0^{(1,2)}$
Nihon
…
Nippon
$s_n^{(1,2)}$

Automatically create synonyms using YAGO*

Ex: $\{s_0^{(0,0)}, s_0^{(1,2)}\}$: **{MIT, Japan}**
"I study at **MIT** and am from **Japan**"
assigned to "Friend 1" of Friends group

Friend 1

C. Fellbaum, "Wordnet," in *Theory and Applications of Ontology*: Computer Applications. Springer Netherlands, 2010, pp. 231–243.

Anonymize sensitive phrases

Anonymize temporal phrases

Conclusion

References

Detect Sensitive Phrases

- **t': Disclosed text**
  - t': I study at MIT and am from Japan.
- Detecting Sensitive Phrases by DCNL*
  - $A = \{a_0, a_1, a_2, ...\}$: set of attributes about a user

Entries in user profiles A		Phrases in blog text t
First name	a_0 = "Adam"	I
Last name	a_1= "Ebert"	…
University	a_2= "**Massachusetts Institute of Technology**"	**MIT**
Nickname	…	**…**
Prefecture	a_n= "Tokyo"	**Japan**

- **Output:** Detect sensitive phrases
  - $P' = \mathscr{D}(A, t') = \{p'_i\} = \{MIT, Japan\}$
- ➔ "**Friend 1**" disclosed information

Friend 1

Evaluation

Anonymize sensitive phrases

Anonymize temporal phrases

Conclusion

References

Number Possible Groups & Friends

- Number of possible groups

$$T = \prod_{i=0}^{N-1} |G_i|$$

  - N: number of sensitive phrases
- Number of possible friends

$$F = \sum_{i=0}^{T-1} \prod_{j=0}^{N-1} |S^{(j, index_{i,j})}|$$

  - $index_{i,j}$: the generalized level for the i-th group of the j-th sensitive phrase

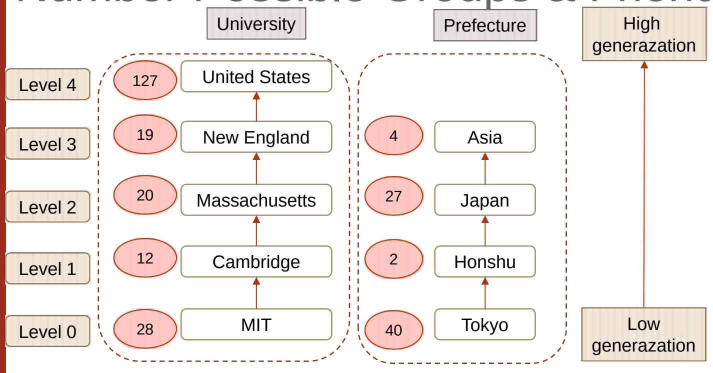
Anonymize sensitive phrases

Anonymize temporal phrases

Conclusion

References

Number Possible Groups & Friends



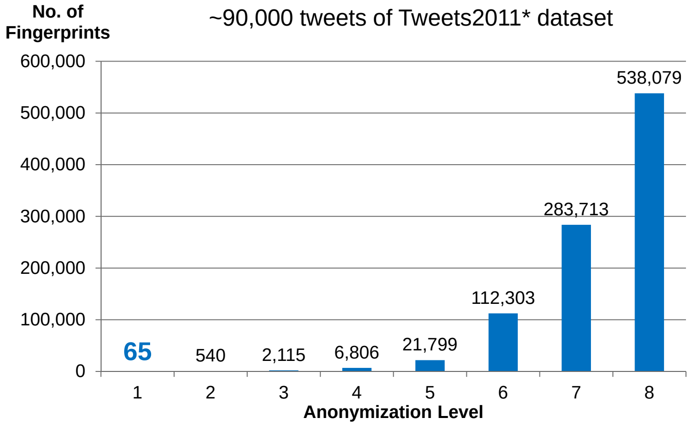
Generalization Schemas

Number of possible groups: 5 × 4 = 20

Number of possible friends: (28 × 40) + (20 × 2) + (28 × 27) + … = 15038

- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

Number of fingerprints



Previous approach [Zheng,2009]: **65 fingerprint/tweet**

Our approach **471.7 fingerprints/tweet**

➔ create enough fingerprints for almost cases on SNS

Anonymize temporal phrases

- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

Privacy on Social Network Services

- Time of user’s activity information is often easily found by crimes
  - Ex: I go out with my family tomorrow.
  - ➔ The crimes enter user’s house at that time

Anonymize temporal information of text to be posed on SNS

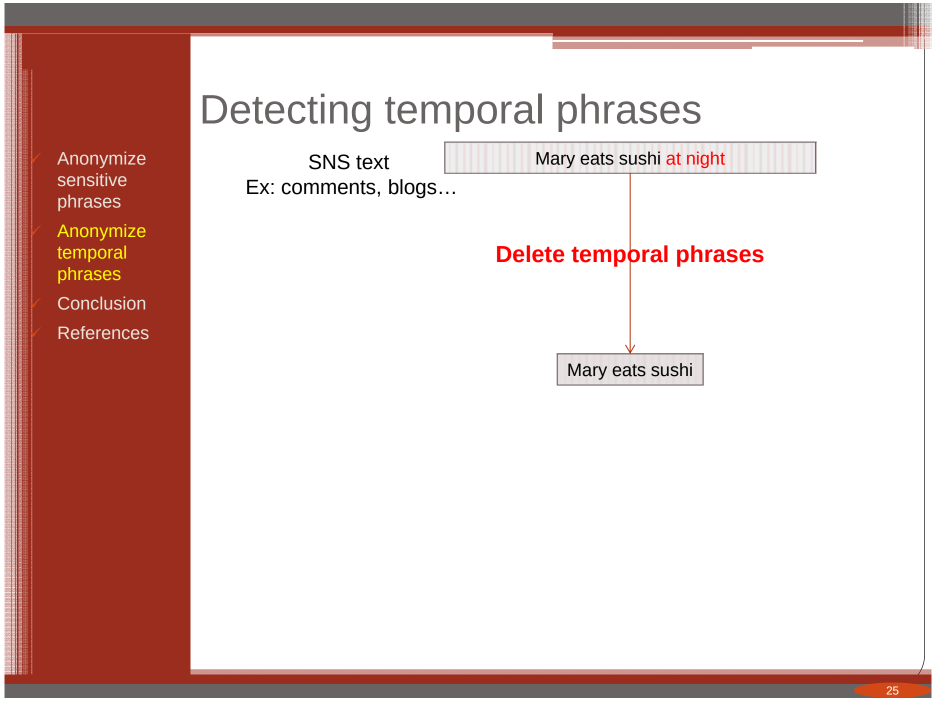
- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

Anonymous temporal phrases

- **Detect** all temporal phrases in texts [Chang,2012]
  - Input: I went to NII **at 9AM**/TIME
  - Output: I went to NII **at**

➔ **Not natural** after removing the detected temporal phrases

Propose deleting all temporal phrases depend on learning structure of parsing tree in a sentence



Detecting temporal phrases

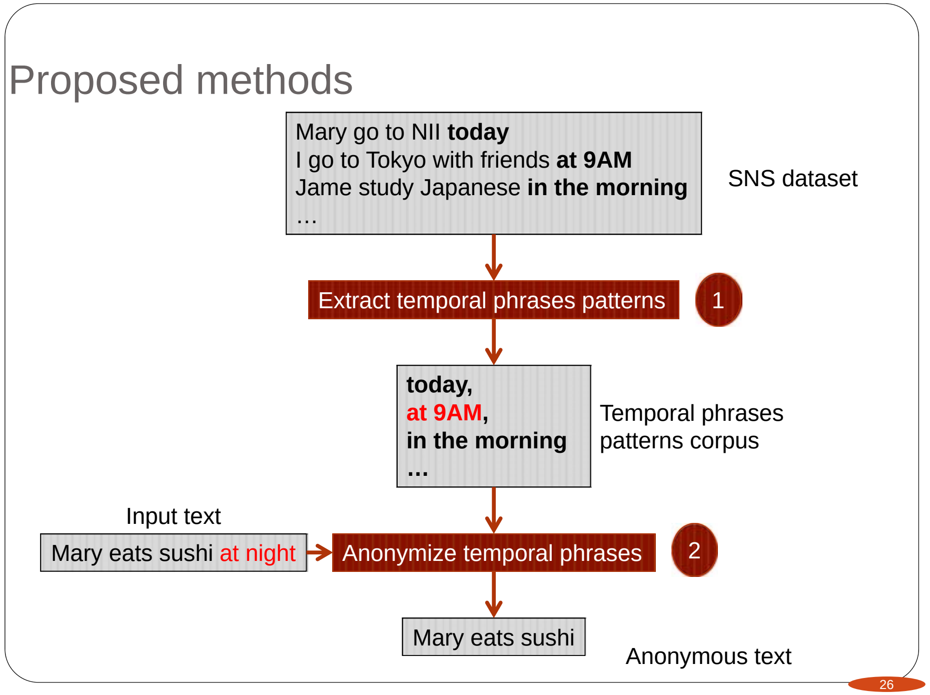
- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

SNS text
Ex: comments, blogs…

Mary eats sushi at night

Delete temporal phrases

Mary eats sushi



Proposed methods

Mary go to NII **today**
I go to Tokyo with friends **at 9AM**
Jame study Japanese **in the morning**
…

SNS dataset

Extract temporal phrases patterns

1

today,
at 9AM,
in the morning
…

Temporal phrases patterns corpus

Input text

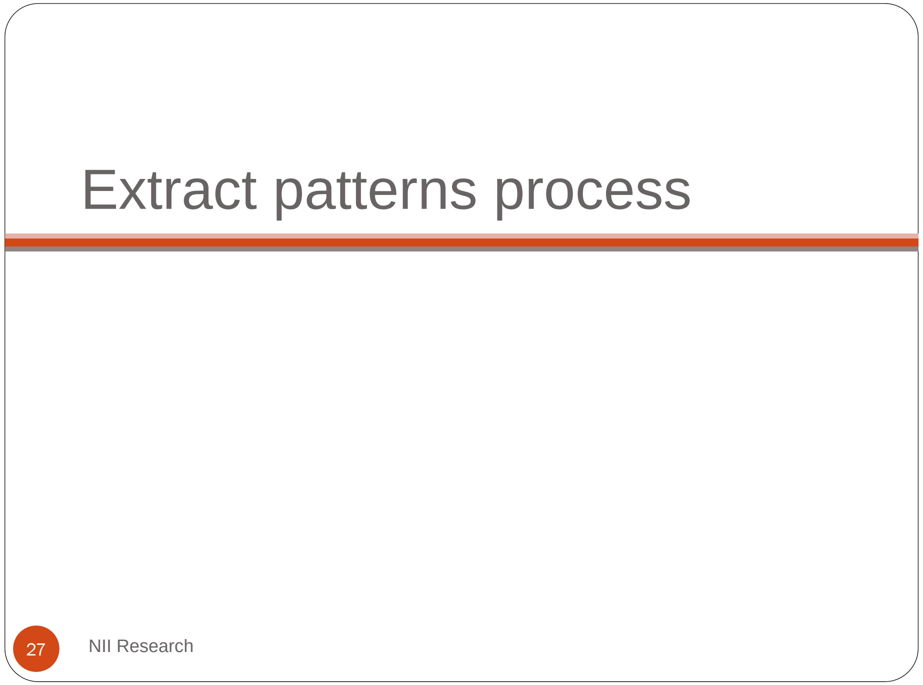
Mary eats sushi at night

Anonymize temporal phrases

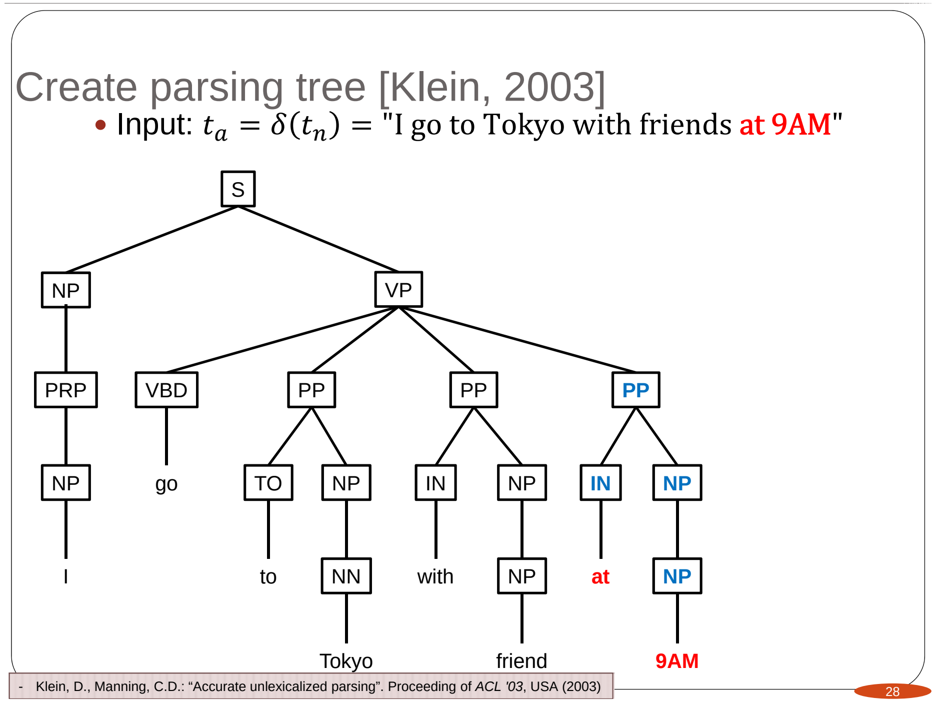
2

Mary eats sushi

Anonymous text



Extract patterns process



Create parsing tree [Klein, 2003]

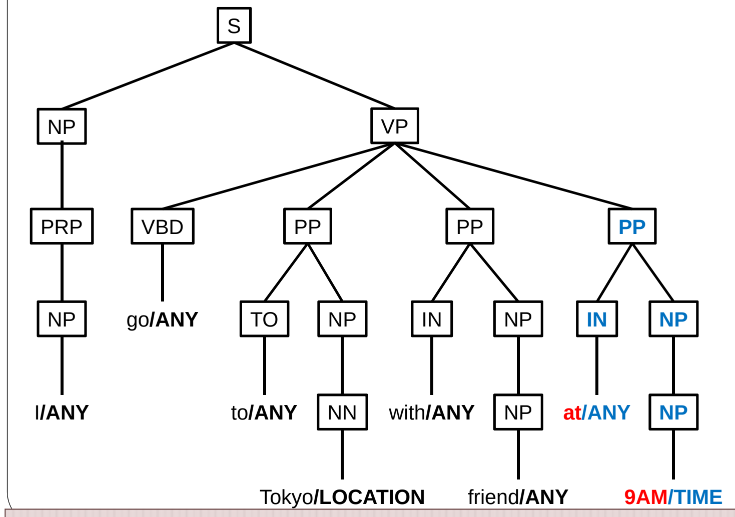
- Input: $t_a = \delta(t_n) =$ "I go to Tokyo with friends at 9AM"

S
NP
VP
PRP
VBD
PP
PP
PP
NP
go
TO
NP
IN
NP
IN
NP
I
to
NN
with
NP
at
NP
Tokyo
friend
9AM

- Klein, D., Manning, C.D.: "Accurate unlexicalized parsing". Proceeding of *ACL '03*, USA (2003)

Annotate temporal phrases [Chang, 2012]

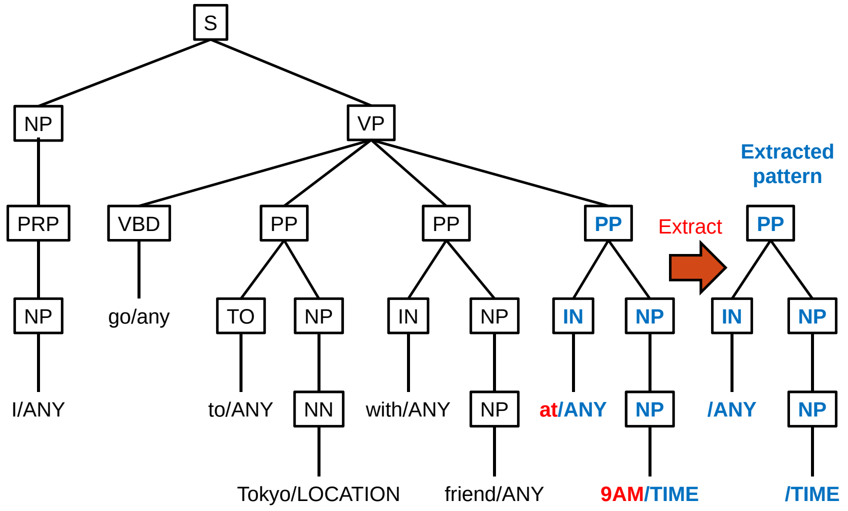
- Input: t_a = "I go to Tokyo with friends at 9AM"



- Chang, A.X., Manning, C.: "Sutime: A library for recognizing and normalizing time expressions". *Proceedings of LREC'12,* Turkey (2012)

Extract patterns

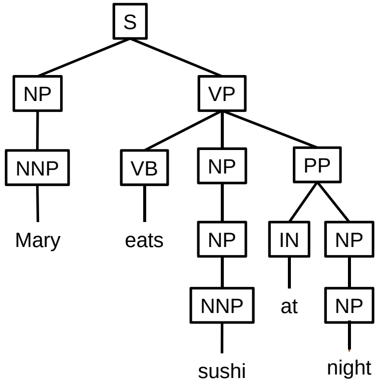
- Input: t_a = "I go to Tokyo with friends at 9AM"



Delete temporal phrases process

Create parsing tree [Klein,2003]

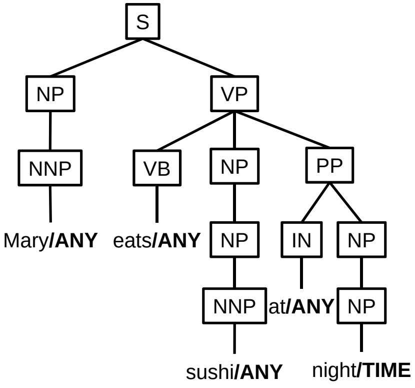
- Input: t'_a = "Mary eats sushi at night"



- Klein, D., Manning, C.D.: "Accurate unlexicalized parsing". Proceeding of *ACL '03*, USA (2003)

Annotate temporal phrases [Chang, 2012]

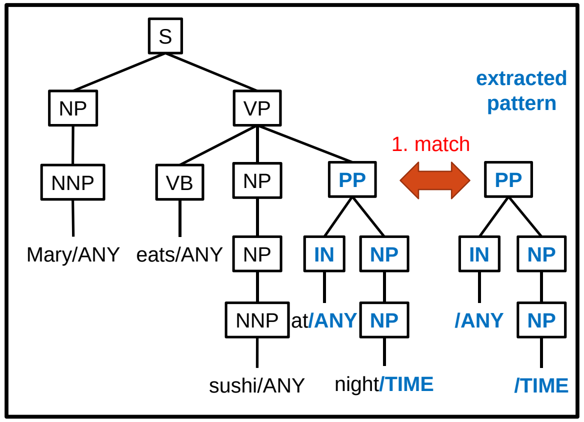
- Input: t'_a = "Mary eats sushi at night"



- Chang, A.X., Manning, C.: "Sutime: A library for recognizing and normalizing time expressions". *Proceedings of LREC'12*, Turkey (2012)

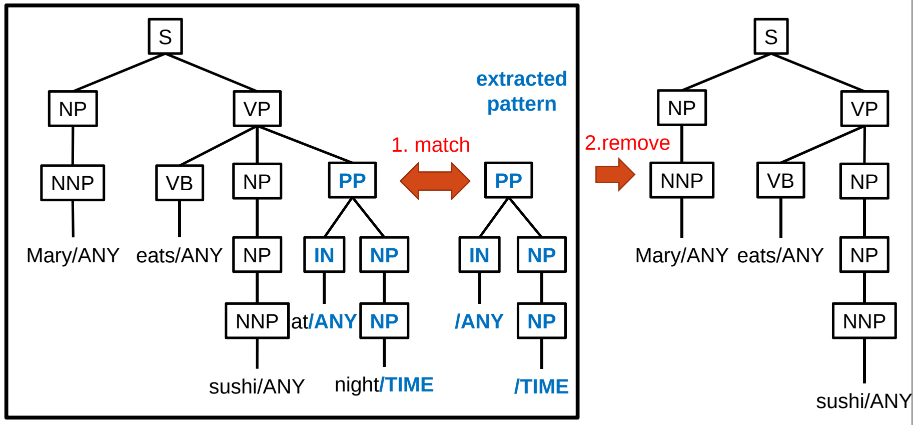
Match with extracted patterns

- Input: t'_a = "Mary eats sushi at night"



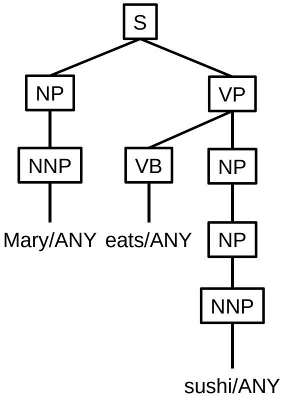
Remove temporal phrases

- Input: t'_a = "Mary eats sushi at night"



Anonymize sensitive phrases
Anonymize temporal phrases
Conclusion
References

Remove temporal phrases



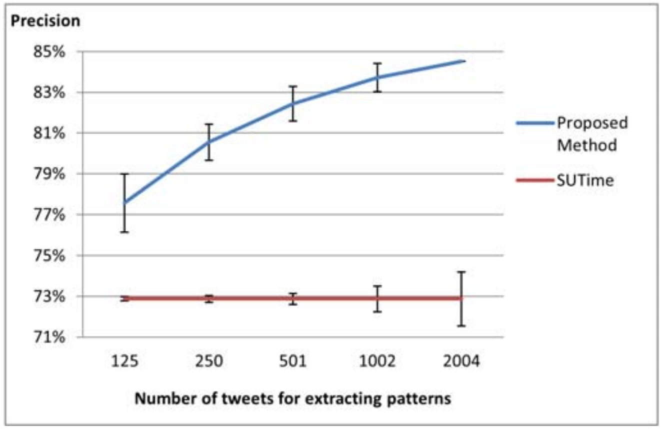
⇒ t'_r = "Mary eats sushi"

Evaluation

- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

Evaluation

- ~2000 tweets of Tweets2011* dataset



* Ounis, I., Macdonald, C., Lin, J., Soboro, I.: Overview of the trec-2011 microblog track. In: Proceeddings of the 20th Text REtrieval Conference (TREC 2011) (2011)

- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

Conclusion

- Addressed problem of information disclosure on social networking services
- Proposed algorithm for automatically creating **anonymous** text to be posted on social networking services by :
  - generalizing sensitive phrases
  - **deleting temporal phrases**
- Future works
  - Anonymous temporal phrases by generalization
  - Anonymous other phrases (location, objectives…)

- Anonymize sensitive phrases
- Anonymize temporal phrases
- Conclusion
- References

References

- H. Kataoka, A. Utsumi, Y. Hirose, and H. Yoshiura. Disclosure control of natural language information to enable secure and enjoyable communication over the internet. In *Security Protocols*, pages 178-188. Springer, 2010.
- L. Sweeney et al. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty Fuzziness and Knowledge Based Systems*, 10(5):557-570, 2002.
- C. Jensen. Fingerprinting text in logical markup languages. *Information Security*, pages 433-445, 2001.
- University prole. http://www3.ibac.co.jp/univ1/mst/info/univinfo_50.jsp.
- C. Fellbaum. Wordnet. In *Theory and Applications of Ontology: Computer Applications*, pages 231-243. Springer Netherlands, 2010.
- I.F. Lam, K.T. Chen, and L.J. Chen. Involuntary information leakage in social network services. *Advances in Information and Computer Security*, pages 167-183, 2008.
- P. Samarati and L. Sweeney. Generalizing data to provide anonymity when disclosing information. In Proceedings of the Acm Sigact Sigmod Sigart Symposium on Principles of Database Systems, volume 17, pages 188-188. Association for Computing Machinery, 1998.
- S. Schrittwieser, P. Kieseberg, I. Echizen, S. Wohlgemuth, N. Sonehara, and E. Weippl. An algorithm for k-anonymity-based fingerprinting. In *10th International Workshop on Digital forensics and Watermarking*, 14 pages. IEEE, 2011.
- S. Gurses. State of the art. *Technology*, 19(1):44, 2011.
- M.K. Arnold, M. Schmucker, and S.D. Wolthusen.*Techniques and applications of digital watermarking and content protection*. Artech House Publishers,2003.
- T. Rethika, I. Prathap, and R. Anitha, “A Novel Approach to Watermark Text Documents Based on Eigen Values,” *Network and Service*, 2009.

Thank you for your attention