

## Anonymizing Textual Data and its Impact on Utility

NSF Grant No. 1012208

Chris Clifton and Luo Si  
*Computer Science*

Karen Chang  
*Nursing*

Victor Raskin  
*Linguistics*

Collaborative with Raquel Hill and Stephanie Sanders (Indiana University)  
and Wei Jiang (Missouri University of Science and Technology)

### *Can we anonymize text?*

- Redaction leaves text unreadable
- Can we use ontologies to desensitize/deidentify text while retaining semantics?

...this indicates that political **philosophy** leads **congressional Republicans** to pay less attention to narrow constituent concerns.

...this indicates that political **belief** leads **congressional humans** to pay less attention to narrow constituent concerns.

### *Is anonymization effective?*

Created (semi-synthetic) dataset and challenge problems

- Can we determine who a record belongs to?
- Can we learn information about an individual?

*Challenging students to “break” different anonymizations*

### *Is anonymized data useful?*

Shadowing Diabetes care study using EHR

*Does analysis of anonymized data give the same results?*

