

# Knowledge Discovery through Text Mining Biomedical Documents Relating Climate and Cancer Incidence

Katherine Gaster  
Wofford College

Research Alliance in Math and Science  
Oak Ridge National Laboratory  
[http://info.ornl.gov/sites/rams2012/k\\_gaster](http://info.ornl.gov/sites/rams2012/k_gaster)

Mentor: Georgia Tourassi, Ph.D.  
Computational Sciences and  
Engineering Division

## Introduction

- Relationship between climate and health widely studied
  - Climatic temperature stress increases cardiovascular disease risk
  - Solar UV radiation increases skin cancer risk
- Cancer incidence rates expected to rise dramatically
- Cancer risk factors: genetic vs. environmental
- Objective: Investigate possible link between climate and cancer incidence
  - How: Knowledge discovery via text mining of biomedical literature
  - Why: Efficient way to assist scientific hypothesis formation

## Methodology

Changing threshold values	<ul style="list-style-type: none"> <li>Values from 0.0 to 1.0; intervals of 0.2</li> <li>Intervals halved three times</li> </ul>
Changing stop word groups	<ul style="list-style-type: none"> <li>Combinations of groups</li> <li>In conjunction with varied threshold values</li> </ul>
Category addition	<ul style="list-style-type: none"> <li>Cancer types; risk factors</li> <li>Used with varied threshold values; stop word groups</li> </ul>

## Study design

- Literature search to find relevant publications
- Text mining using Piranha software
- Investigate impact of Piranha's key features

## Results

### Discovered **expected** associations

- Skin cancer risk and UV exposure
- Breast cancer / prostate cancer risk and genetic profile

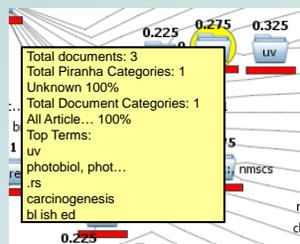


Figure 1. Expected association: cancer risk and UV exposure.

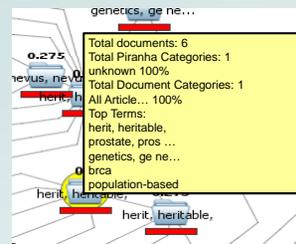


Figure 2. Expected association: breast or prostate cancer risk and genetic profile.

### Discovered **unexpected** association: increased skin cancer risk in kidney transplant recipients

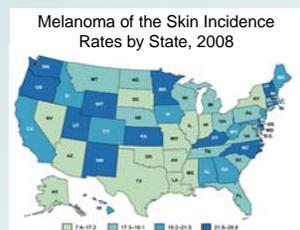


Figure 3. Map depicting relative melanoma incidence for each U.S. state in 2008.

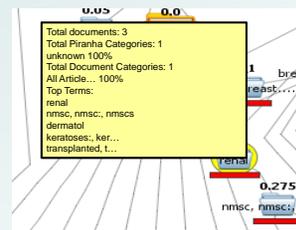


Figure 4. Unexpected association: kidney transplants and skin cancer.

- Obtained additional publications to investigate association further
  - Completed literature search for publications referring to kidney transplants and skin cancer
  - Analyzed supplementary documents with and without original documents

## Conclusions

- Text mining revealed expected associations between climatic factors and cancer types
- Unexpected association discovered: increased skin cancer risk in kidney transplant recipients
- Supplementary articles clustered as expected

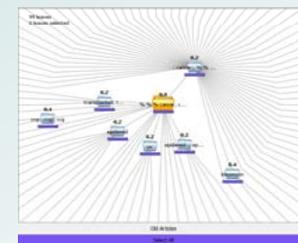


Figure 5. Document clustering of original articles. Original threshold values, no stop words.

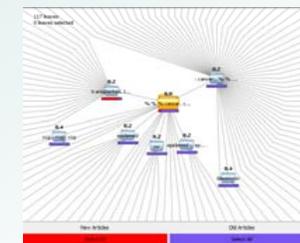


Figure 6. Document clustering of original and supplementary articles. Original threshold values, no stop words.

## Future Work

- Implement investigation linking actual transplant records and cancer incidence data
- Determine if higher association exists between skin cancer and kidney transplants in sunny climates
- Generate new hypotheses that may be tested with traditional, long term epidemiological studies
- Develop model to predict future expected cancer incidence based on environmental factors

- Crandall, C.G., & Gonzalez-Alonso, J. (2010). Cardiovascular Function in the Heat-Stressed Human. *Acta Physiologica*, 407-23.
- Jemal, A., Center, M. M., & DeSantis, C. (2010). Global Patterns of Cancer Incidence Rates and Mortality Rates and Trends. *Cancer Epidemiology Biomarkers & Prevention*, 1893-907.
- U.S. cancer statistics: an interactive atlas [Image]. (n.d.). Retrieved from [http://apps.nccd.cdc.gov/DCPC\\_INCA/DCPC\\_INCA.aspx](http://apps.nccd.cdc.gov/DCPC_INCA/DCPC_INCA.aspx)
- van der Leun, J. C., & de Grujij, F. R. (2002). Climate change and skin cancer [Abstract]. *Photochemical and Photobiological Sciences*, 324-26.

