



Regional Level Community Engagement – A collaborative e-democracy through the use of technology driven Visualisation and Advanced Computational Tools

Mr. Wayne Delaforce - QUT (Previously)

Assoc. Prof. Nasir Butrous - ACU

Dr. Ann Bramwell - ACU

Mr Tim Davis - ACU

Ms Bernadette Savage – QUT

Sponsored by QCIF



Project Summary

- Designed to further enhance and extend a community based e-democracy/e-participation vision and model.
- Used intelligent tools to distil and compare thematic current issues reported in three local community newspapers over three and a half years.
- Represents collaborative regional level research project that reflects the benefits of higher education participating with the community to create and share new knowledge.
- The project represents the collective work of staff from three universities across a number of discipline boundaries.



Research Methodology

- Quest Community papers – North Brisbane: City North News, Northside Chronicle, and North-West News.
- Historical Text-based story data from each newspaper for 2004, 2005 and 2006 was analysed for themes and content using computer software tools.
- A visual audit was done using six months of 2007 data for comparisons using different classification methods.



Catchment Areas of QUEST COMMUNITY NEWSPAPERS

NorthWest News

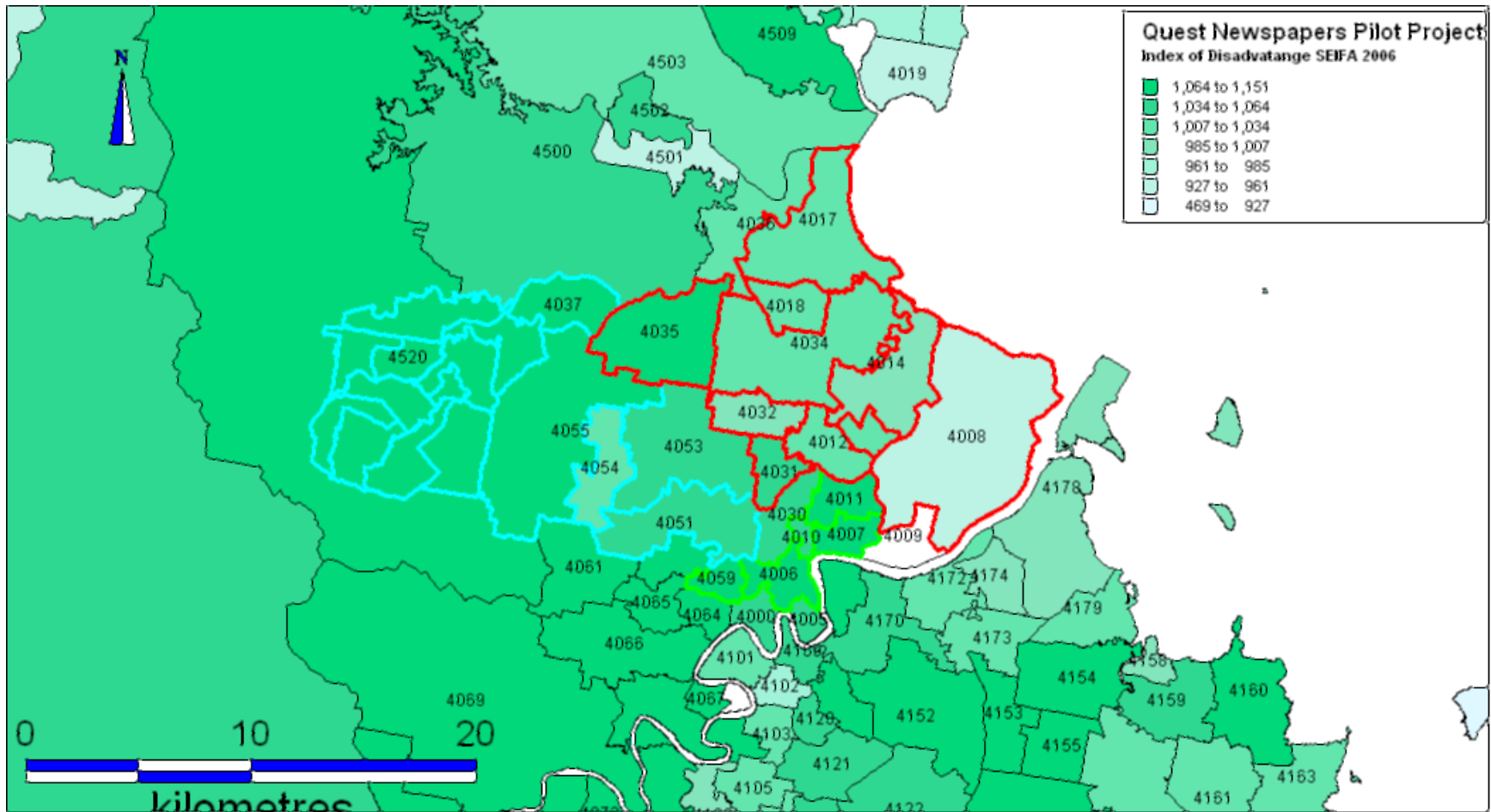


City North News

Northside Chronicle



Population Density of Quest Papers



The Pilot Project

- an application model that uses software tools to explore large datasets for thematic issues.
- continuous cycle of communication between various players within the three spheres: **Society** – citizens, lobbies and opinion leaders; **Media** – multimedia, agencies, market researchers; and **Politics** – administration, legislation and advisors.
- Technology can provide a public space necessary for individuals to engage in public deliberation over normative values.
- the opportunity to comprehend an enriched conception of citizenship through which to construct the possibility for participation in democratic self-government.



Data Analysis

- This pilot is studying the interactions between society and the media using a small group of community newspapers covering the northern suburbs of Brisbane.
- Analysis of Quest data from three Brisbane North community newspapers was done using a variety of methods, C code on the supercomputer, SAS TextMiner, Leximancer, and Wordstat.
- All of the programs use a variety of simple and complex methods to extract text components and meaning from the text.



Data Analysis

- Some of the methods used in the programs are listed below:
- **Corpus and Computational Linguistics:** a word can be characterised by the words that tend to appear near it, and not apart from it.
- **Machine Learning:** Leximancer uses a machine learning optimisation approach for iteratively growing a thesaurus of words around a set of initial seed words.
- **Complex Networks Theory:** The cluster map presentation is heavily influenced by the area of Complex Networks,
- **Physic:** The idea of a measurable short-range order between words was influenced by solid-state physics
- **Content Analysis:** The quantification of knowledge within text by means of coding or tagging of text segments using a set of concepts
- **Information Science:** The principles and guidelines for indexing and navigating large amounts of information in a complex concept space



Data Analysis

- Statistical text analysis techniques such as Latent Semantic Analysis (LSA) and the Hyperspace Analog to Language (HAL) have demonstrated that highly useful and reliable information can be extracted from the word co-occurrence information in text.
- Even though each software package produced a degree of useful results such as themes, word counts, key words and phrases there was still a degree of expert knowledge required to train the software to recognise what was meaningful and contextually correct.



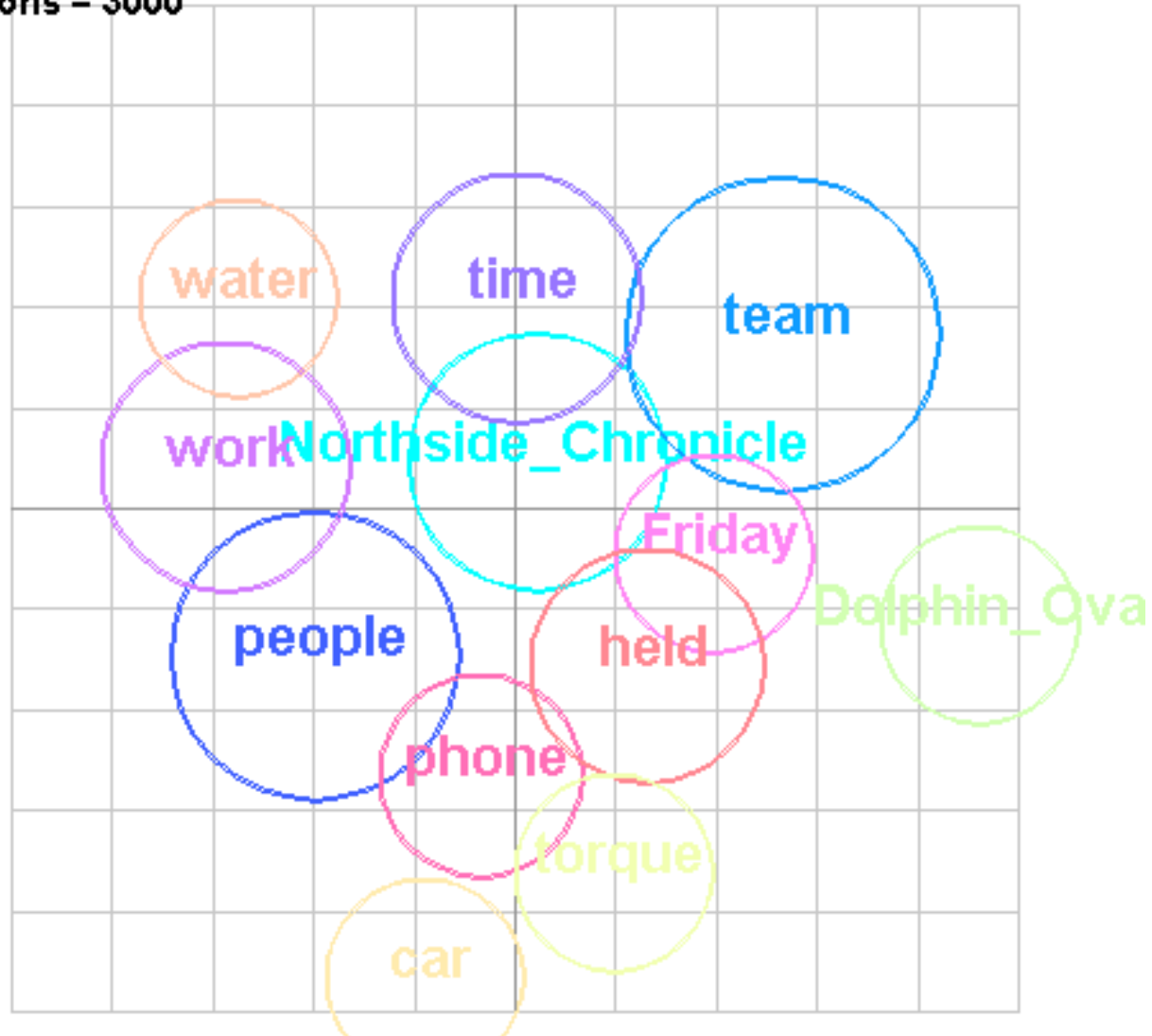
Major Findings: City North News 2004-2006

Iterations = 1000



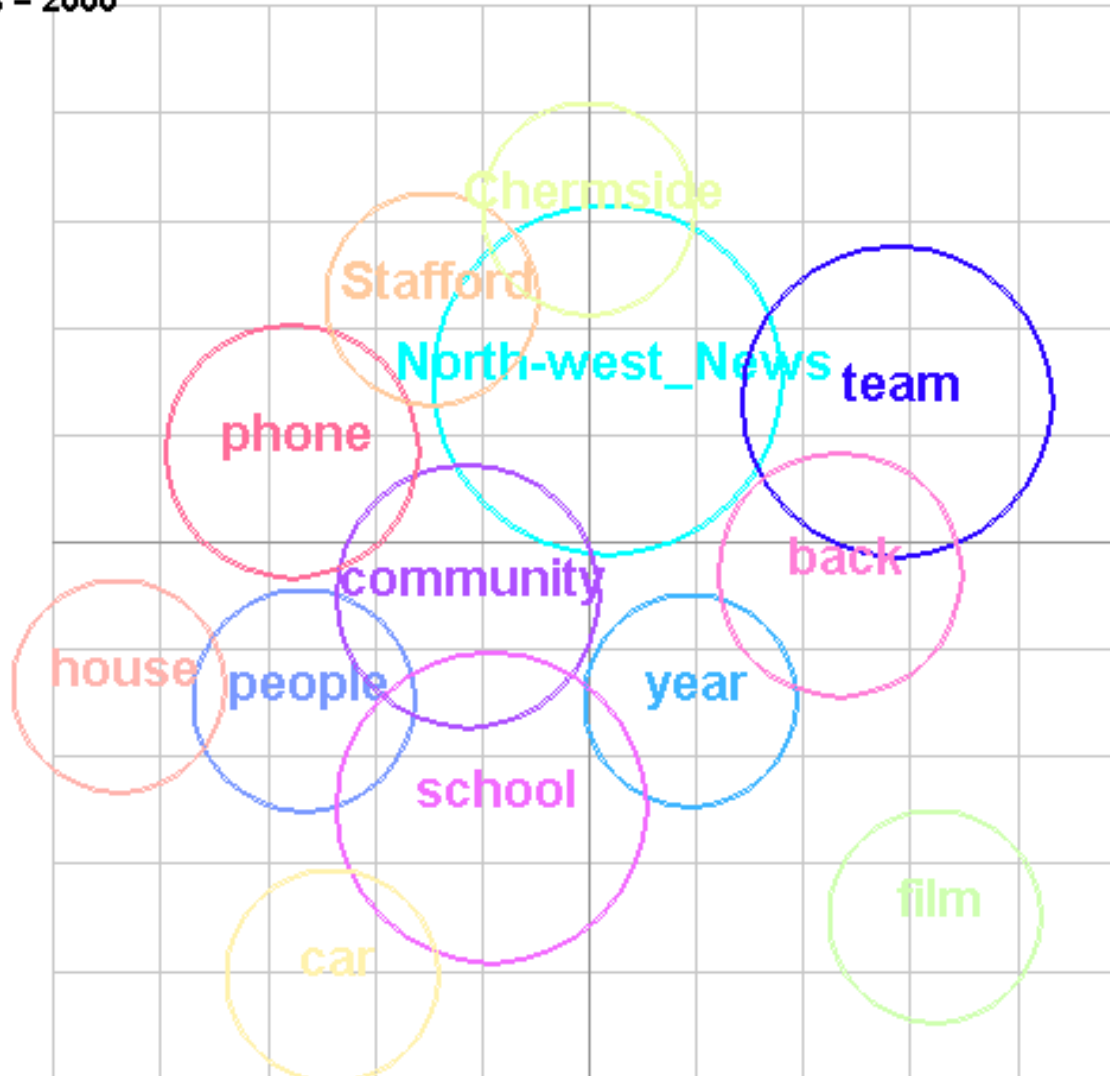
Major Findings: Northside Chronicle 2004-2006

Iterations = 3000



Major Findings: North West News 2004-2006

Iterations = 2000

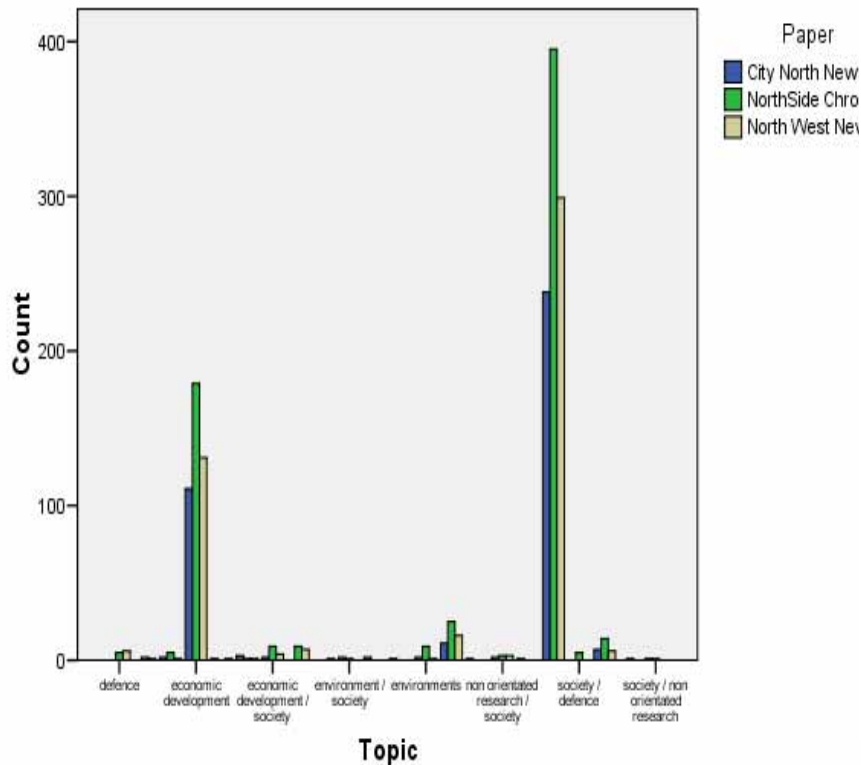


Major Themes

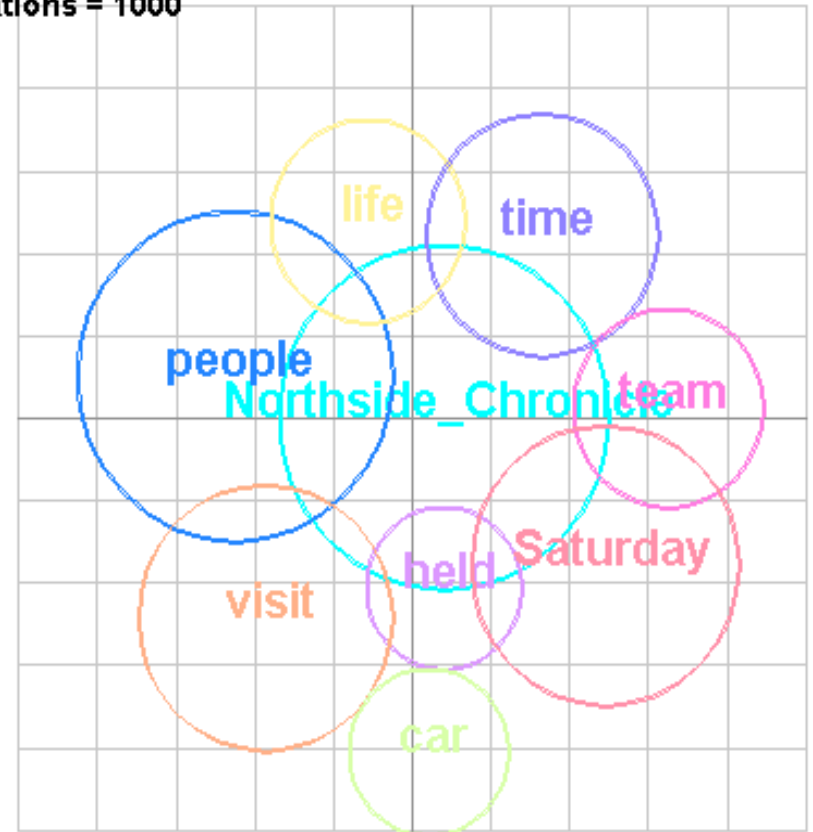
City North News	%	Northside Cronical	%	North-West News	%
People	33%	People	29.4%	People	26.3%
Win	21%	Team	20%	Time	20.7%
Team	19.9%	Children	17.8%	Home	16.3%
Home	19.9%	Community	17.6%	Children	16.2%
Day	19.2%	Saturday	17%	School	16.1%
Work	19.1%	Home	16.6%	Team	14.9%
Brisbane	18.7%	Win	16.3%	Community	14.7%
Bedrooms	18.2%	Work	15.7%	Work	14.4%
Back	16.3%	Brisbane	15.5%	Saturday	13.9%
House	15.3%	School	15.5%	Brisbane	13.8%

Visual audit using six months of 2007

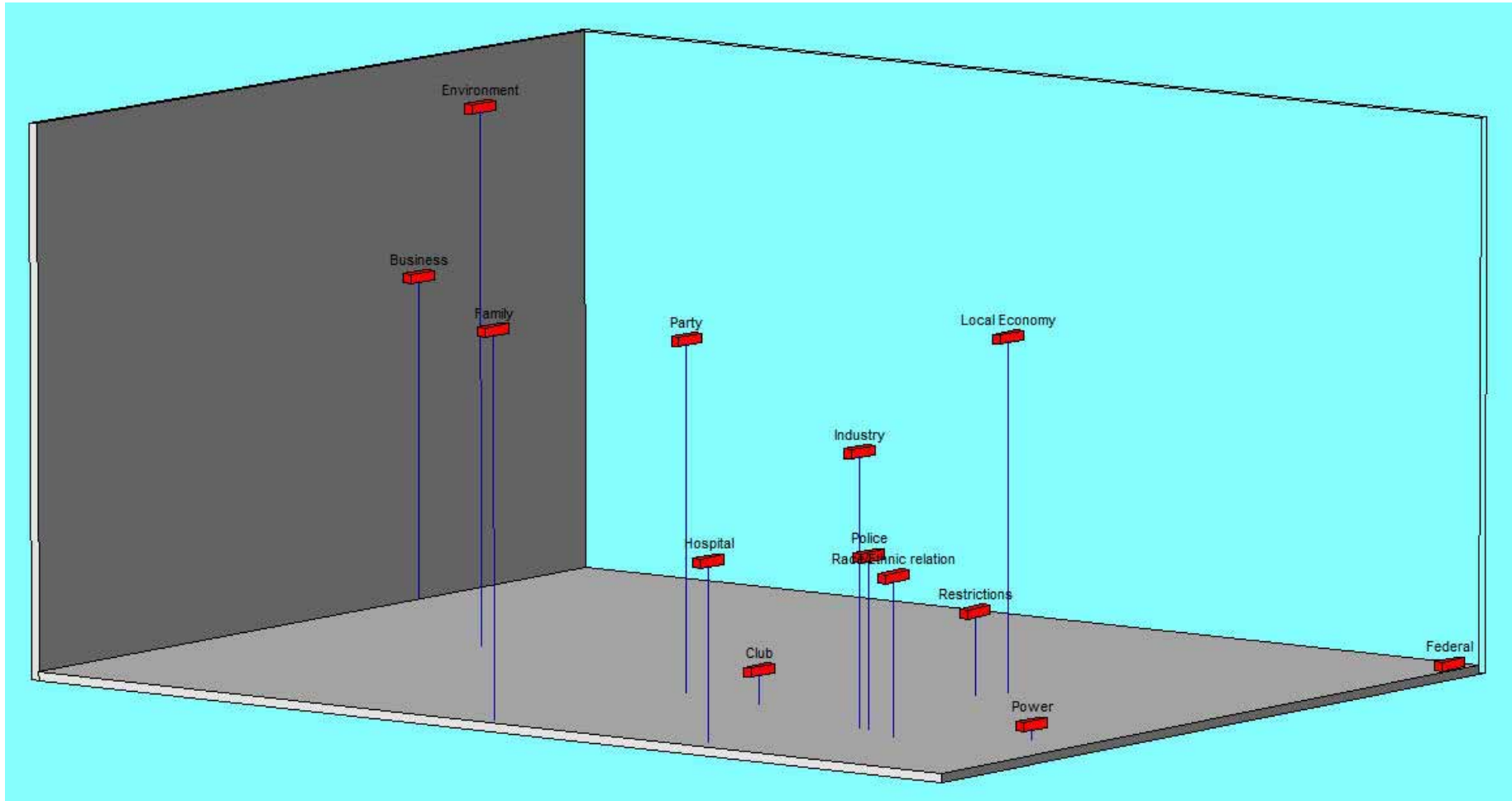
Bar Chart



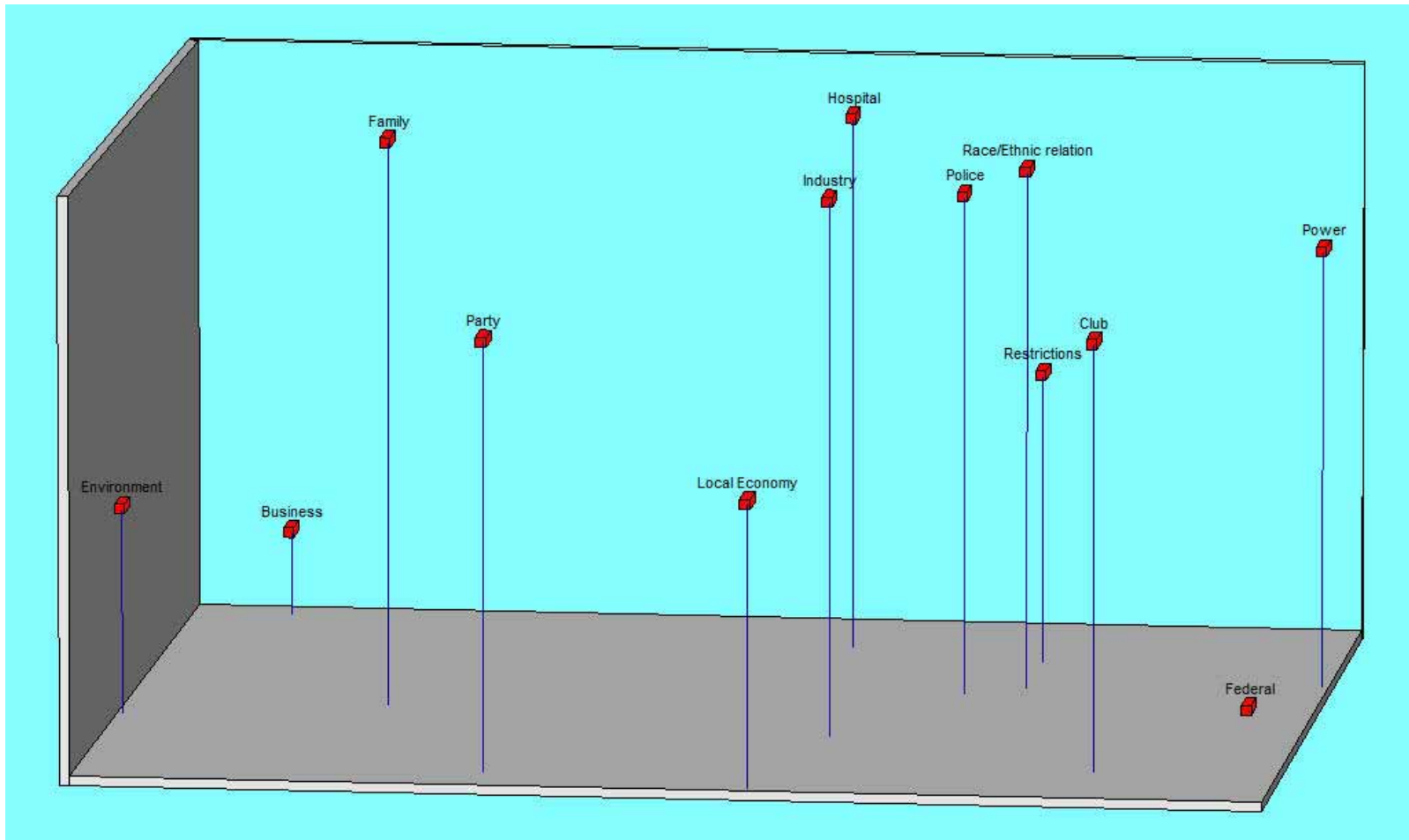
Iterations = 1000



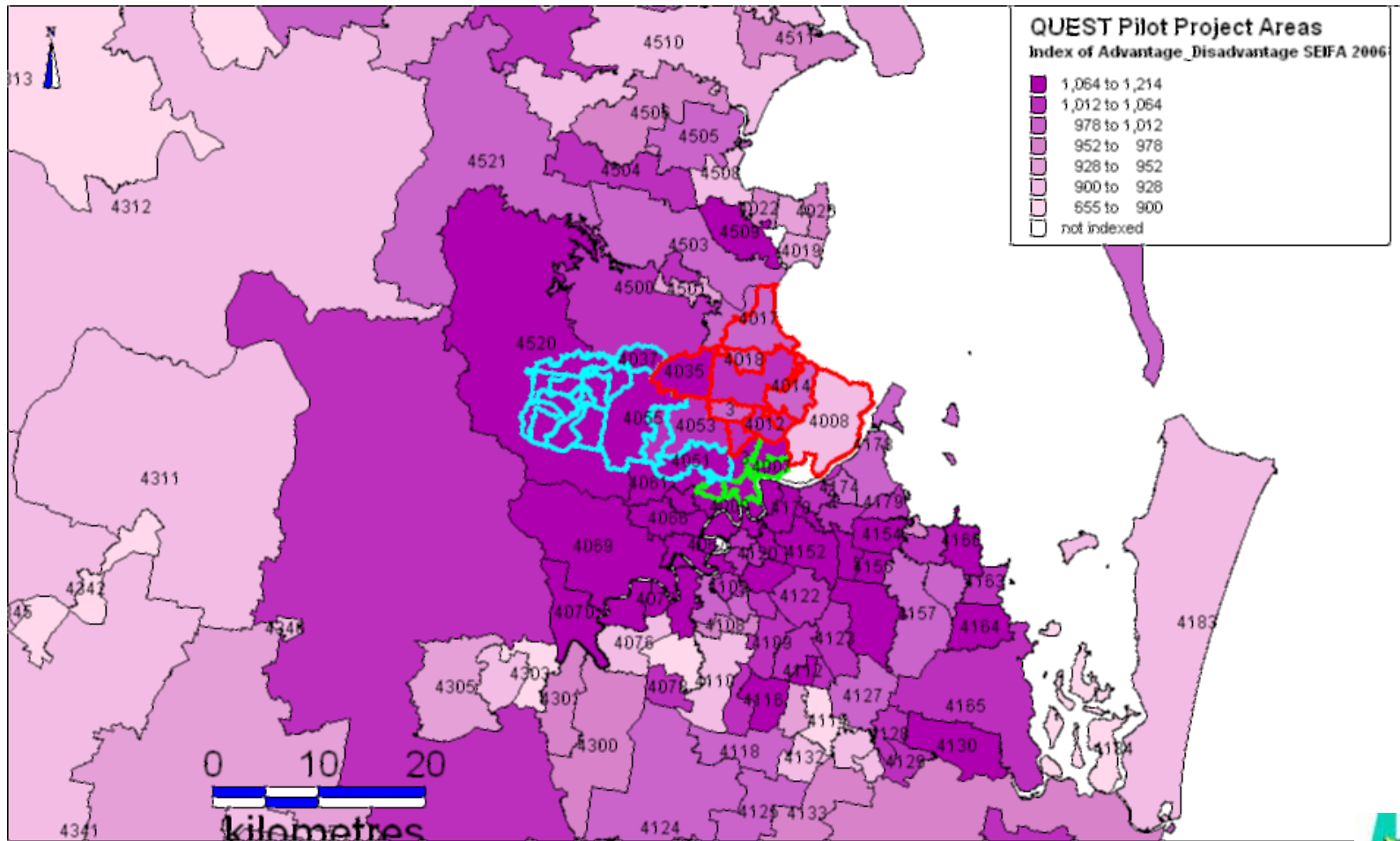
Sorensen's Coefficient (used to compare similarity of sample sets)



Jaccard's Coefficient (used to compare similarity and difference of sample sets)



Index of Advantage - Disadvantage



Conclusion

- the language used in the community newspapers was simple and targeted to the geographic and demographic local areas very colloquial and context based.
- There is a marked difference in the types of words used to classify topics using the ABS classification which is disjunct and clinical, for a machine learned program to be able to work with the language used by these classification systems many hours of teaching for context relationships and key words would be required.
- The level of human and artificial intelligence become intertwined and very interesting.
- The use of underpinning mapping data and thematic visualisation as images allowed better understanding of the data of each stage.
- The supercomputers and complex software programs easily handled large volumes of textual based data; the human interface for interpretation however is still useful where complex language issues arise. A study using a larger geographical area and larger text data files would be possible.

Further Research

- In order to develop an intelligent system with prediction capabilities from this pilot it will be important to follow-up how issues gain momentum with people in the community in real time.
- This would be part of the next stage in the project utilising the real-time data from www.whereilive.com.au web site to explore the democratic process played out by the media in the online forums.
- Further work would be done using spatial comparisons and analysis using postcode variables with Geographic Information Systems technologies to enhance any findings.
- The knowledge gained in this pilot will inform the next series of questions to find the links between media, community and government