MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

A Text Mining Approach to Explain What Being Actively Engaged in Life Means

Richi Nayak[a], Laurie Buys[b]

[a] *Faculty of Information Technology,* [b] *Centre for Social Change Research Queensland University of Technology, Brisbane, Australia*

Abstract

This paper shows the findings and analysis of a national survey data on 'Active Ageing'. We examine the applicants responses on "what being actively engaged in life means to you?" with the text mining approach. The analysis shows that the 'Family', 'Friend' and 'Activity' were the most important concepts appearing across all age groups and both genders.

Keywords:

text mining, data mining, active ageing

Introduction

It is widely accepted that many first world countries are experiencing extended life expectancies and reduced fertility rates, which is resulting in an ageing population. A definition of active ageing is "the process of optimizing opportunities for physical, social, and mental well-being throughout life in order to extend healthy life expectancy [2]. The Australian Active Ageing (Triple A) study at the Queensland University of Technology (QUT) has conducted a national-wide postal survey to collect the responses of older people on a wide range of questions related to 'work', 'learning', 'social', 'spiritual', 'emotional', 'health and home', 'life events' and 'demographics'. The survey also included an open-ended question that the applicants responded to was "Briefly describe what being actively engaged in life means to you?". The purpose of this paper is to present the findings and analysis of the open-ended question in the survey using the text mining techniques [1]. This paper attempts to gain a better understanding of the impact of aging upon older Australians' engagement in society and their sense of well being. The findings and analysis are based on the responses from 2645 (45% response rate) participants over the age of 50 years residing across Australia.

Method: Text mining

Text mining extracts the important concepts and emerging themes from the collection of text sources using computation linguistics, machine learning and information science methods [1,3]. We utilize the Leximancer text mining application [3] to extract relationships from the text data. It is based on the bayesian theory using the words which make up a sentence to predict the concepts being discussed. Each *term* appearing in the text data is analyzed to form a *concept* and then many concepts are put together to form a *theme* based on the co-occurrence of terms and concepts in the text.

Data

The Triple A study data includes responses of 2645 members of a large Australia wide seniors organization on a wide range of questions related to their life. The survey includes 178 questions grouped into 8 sections covering the life aspects of social, health and home, emotional etc and an open-ended question. The sample included a reasonable balance of gender (57% female) and state representation (NSW: 32%, Vic: 24% and Qld: 20%). The respondents were quite comfortable financially, well educated and majority were living with a companion, in their own home in a metropolitan area.

Table 1 - Categorization by Age and Gender

Grouping	Category	Qty
All comments		1964
Grouped by age	55-64 65-74 75 & over	1252 474 238
Grouped by gender	Male Female	870 1094
Males grouped by age	55-64 65-74 75 & over	523 230 117
Females grouped by age	55-64 65-74 75 & over	729 244 121

Process

Preprocessing such as (1) removal of surveys (instances) that did not have sufficient information of age, gender etc, (2) application of spelling and grammar checking function in Excel in order to locate and correct the misspell words, incorrect abbreviations and slang, and (3) stop word removal and stemming were performed. The data file was then arranged and split into categories of both Age and Gender that allowed us to thoroughly compare and contrast the different combinations of categories (Table1).

Results and discussion

Comparison by age groups

Age group 55 – 64: The most significant concept and theme is 'Family'. 'Life' and 'Active' are next in the level of frequency. There is an integral relationship between 'family' and 'time' as well as having close ties with the 'enjoy' entity. 'Work' is considered a significant concept ranking highly for both males and females. 'Work' and 'Active' aspects comes to the fore in Males whereas 'Health' and 'Social' aspects are more frequent in the Females. This may indicate that Males place a greater importance in physically active aspects while Females signal social themes as figuring in their perception of active ageing.

Age group 65 – 74: This age group shows that they consider an active life to being engaged in it, living a full life, having a life with purpose and friendship in terms such as visit, involvement and participation. In this age group, males considered 'work' a more important concept than females. Both males and females emphasize on Family. However the terms that form the 'Family' concept in both groups differ. The females were concerned with interaction with emphasis on meeting, touching, visiting, whereas the males, although the terms indicate a need to be interactive, are less personal. Their terms include social, involved, community, and activities. Although it is a subtle difference, it could possibly indicate that for males, family is also a source of activity or social events. Whereas with the females, it could be perceived as a more emotional connection as opposed to as a source of activities or social connectivity.

Age group 75+: Ageing is perceived as having a close association with family and friends and maintaining an engagement in life through activities with these people. The concept of 'health' emerges as a significant concept, more highly ranked by the males than the females. This differs from the prior age groups indicating that this group may be considering health issues more frequently in relation to active ageing.

A very interesting difference that appears between male and female of this age group is the 'enjoy' concept. Female associate enjoy with life concept. However, male segregate the enjoy concept from either of family or active concept. Male population subtly emphasizes more on 'active' concept in comparison to female. Male population being actively means being involved with people. The female population seems to relate being actively with being involved with their loved ones, friends and social groups such as church.

Comparison by gender

All Females: Ageing is perceived as having a close association with family and friends and enjoying the activities with these people. Ageing is also perceived as being active in life through the time based work activities.

All Male: Ageing is perceived as being active in life and spending time with family and friends.

The most significant difference between the results of the males and females, are that where the females population were concentrated around 'Life', 'Family' and 'Friends', the male population seems to include a wider range of concepts including 'Active' and 'Work' as heavily weighted.

Using the whole population

To understand the behavior of the whole population including various age groups and all gender, we grouped the respondents according to the way they reflect active ageing using the clustering method [4]. This method also utilizes the computation linguistics to analyses the terms that mainly form a theme and consequently puts them into the same group. All respondents are grouped into 6 clusters. Following is the detail of each cluster including the few top-most terms due to which the respondents were put together in same group along with number of similar respondents in that cluster according to the written response.

Cluster 1: people, thing, good, new, health (1073)

Cluster 2: friend, social, work, child, grandchild (736)

Cluster 3: physical, mental, good, health, activity (175)

Cluster 4: mental, physical, happy, independent, family (136)

Cluster 5: help, other, helping, helping others (99)

Cluster 6: body, mind, help, good (89)

Cluster 1 is the biggest cluster in the dataset, which gives a general description for the meaning of active aging. The concepts covered are as follows: learning new things and skills, maintain good health, dealing with other people etc. Cluster 2 focuses more on the aspect of personal life, involves the family life, the friendship and the attitude to society. These respondents would like to enjoy with their family, look after their grandchild, stay with their friends, participate in social event and contribute to the society. Respondents in cluster 3 realize good health (mental and physical both) is the basis for happy life and are willing to participate in activities according to their health status. Cluster 4 respondents find the mental and physical health along with the family surroundings as key aspects for active ageing. People in cluster 5 wish to find out their personal values by helping others, in the aspect of daily life or whatever and wherever others need. The responders in cluster 6 wish to be interactive with other people to make their body and mind active as well as help the others.

These clusters show the various characteristic that each group of common respondents have. Further drilling to each group will identify more common concepts and themes.

Conclusion

A strong common themes across all groups, male, female and age groups were that 'Family', 'Friends' and 'Activities'. The younger age group considers 'Work' an important variable in active ageing, and this drops off to be important for only males in the next group, and is not ranked in the 75 and older group. Within the eldest group however, 'Health' emerges as an important concept, and given that this group is in a minority in the population as a whole, could explain that lack of representation overall of health concerns. These trends seem to indicate that the respondents are considering factors that are relevant to them in the immediate timeframe when responding to the question without consideration toward their future. There are various areas where these results may have impact and warrant some consideration such as future planning for government funding of community activities.

References

[1] Hearst, M. A. (1999): *Untangling Text Data Mining*. The 37th Annual Meeting of the Association for Computational Linguistics, Maryland, June 20-26, (invited paper).

[2] Kalache, A. 1999. *Active Ageing Makes the difference.* World Health Organization. Bulletin of the World Health Organization. 77 (4): 299.

[3] Leximancer (Accessed September 28th, 2006). http://www.leximancer.com/cms//index.php?option=com_content&task=view&id=59&Itemid=114

[4] SAS Text Miner. (Accessed September 8th, 2006) http://www.sas.com/technologies/analytics/datamining/textmine

Address for correspondence

Richi Nayak, School of Information Systems, FIT, QUT, GP, GPO Box 2434, Brisbane, QLD 4001, Australia.
Email: r.nayak@qut.edu.au
Phone: +61 7 3138 1976 Fax: +61 7 3138 1214