International Conference on Advanced Information Engineering and Education Science (ICAIEES 2013)

Development and Application of a Chinese Webpage Suicide Information Mining System (SIMS)

Penglai Chen, Jing Chai, Lu Zhang, Debin Wang
School of Health Service Management, Anhui Medical University, Hefei 230032, Anhui, China
dbwang@vip.sina.com

Abstract - *Background:* Suicide is becoming a major public health issue worldwide and in China. With growing popularity of the internet as well as participation of ordinary people in posting and transmitting messages on the internet, webpage coverage and its impact of reported suicide events has been and will continue to increase rapidly.

Objectives: This study aims at designing and piloting a convenient Chinese webpage suicide information mining system (SIMS) to help search and filter required data from the internet and discover potential features, trends and causality of suicide. *Methods*: SIMS utilizes Microsoft Visual Studio 2008 as development platform, SQL 2008 as database manager, and C# as programming language. It collects raw Chinese webpage data via popular search engines; cleans the raw data via completeness, duplication and relevance checks using trained models, string or sentence comparison plus minimum manual help; translates the cleaned texts into quantitative data through models and supervised fuzzy recognition; and analyzes and visualizes suicide related variables by self-programmed algorithms. *Results:* The SIMS developed comprises five main functions i.e., suicide news collection, suicide blogs collection, data filtering and cleaning, data extraction and translation, blog suicide ideation estimation and data analysis and presentation. Data collection provides a user-friendly interface for retrieving suicide-related news and blogs from Chinese webpages and downloading them into SIMS database. Data filtering and cleaning performs completeness, duplication and relevance checks of the data gathered. Data extraction and translation facilitates deriving structured data about suicide from unstructured texts. Suicide ideation estimation assigns values of the suicide risks to blogs under concern using default or user built models. SIMS produces a set of useful indicators including frequencies and compositions of webpage reported suicide events and blogs with high suicidal ideations published in total and by different genders, ages, regions, causes, methods, years, months and hours etc. presented in easily understandable diagrams. *Conclusions:* SIMS provides a novel and practical means for monitoring and understanding suicide. It proposes useful aspects as well as tools for analyzing the features and trends of suicide using data derived from the internet as supplements to traditional suicide reporting and epidemiology surveys. Although SIMS was designed specifically for suicide, the overall architecture, strategies and techniques can be easily adapted to similar systems for addressing other diseases or health problems.

Index term - suicide, news, blogs, data mining, support system

1. Introduction

Suicide is a world-wide major public health issue [1-2]. In China, the number of deaths caused by suicide ranked fifth among all deaths and according to a World Health Organization (WHO) report, the epidemic was estimated as the fourth-leading cause for China's disability-adjusted life year losses [3-6]. Suicide events not only do great harm to the victims themselves and their families, but also arouse strong concerns to the society, catching wide spread media (including web media) attention and coverage. With the growing popularity of the internet as well as participation of ordinary people in posting and transmitting internet messages, webpage coverage and its impact of reported suicide events has been and will continue to increase rapidly. Webpage-reported suicide news can not only serve as a warning to the general public, but also provide supplementary data for suicide monitoring, analysis and prevention [7-8].

Blogs are another fast growing form of web-based information and blogs regarding suicidal ideation have special significance for understanding and controlling suicide events [9]. The descriptions and comments of suicide antecedents, processes, consequences, experiences, perceptions etc. by the very bloggers who have experienced the suicide event of either closest friends or themselves, may contain important information and clues [10]. For many individuals with suicidal ideation, they would rather tell their true thoughts in a virtual world than a real one [11]. This provides a unique and convenient way for understanding suicide ideations and targeting high risk individuals.

Given the exponentially growing news, blogs etc., extracting required suicide-related information from vast amount of unstructured text data poses a huge challenge. Fortunately, there have emerged a large number of powerful search engines and data mining techniques [12-13]. This study aims at taking the advantage of these technologies and resources and designing a convenient Chinese webpage suicide information mining system (SIMS) to help search and filter required information from the internet and discover potential features, trends and causality of suicide. According to our systematic literature search, although a few researchers have reported close relationships between the frequencies of searches of suicide-related terms through popular search engines and reported suicide prevalence rates [14,15], literature on monitoring and analyzing suicide-related news, blogs and other formats of data on the internet are scarce.

2. Methodology

A. Programming Tools and Strategies

SIMS utilized Microsoft Visual Studio 2008 as development platform, SQL 2008 as database manager, and C# as programming language. It adopted "rapid prototyping", a common strategy in the field of software engineering which comprised five steps: (a) quick browsing of suicide-related web pages by participating researchers; (b) concept definition via discussions between the researchers; (c) implementation of

© 2013. The authors - Published by Atlantis Press

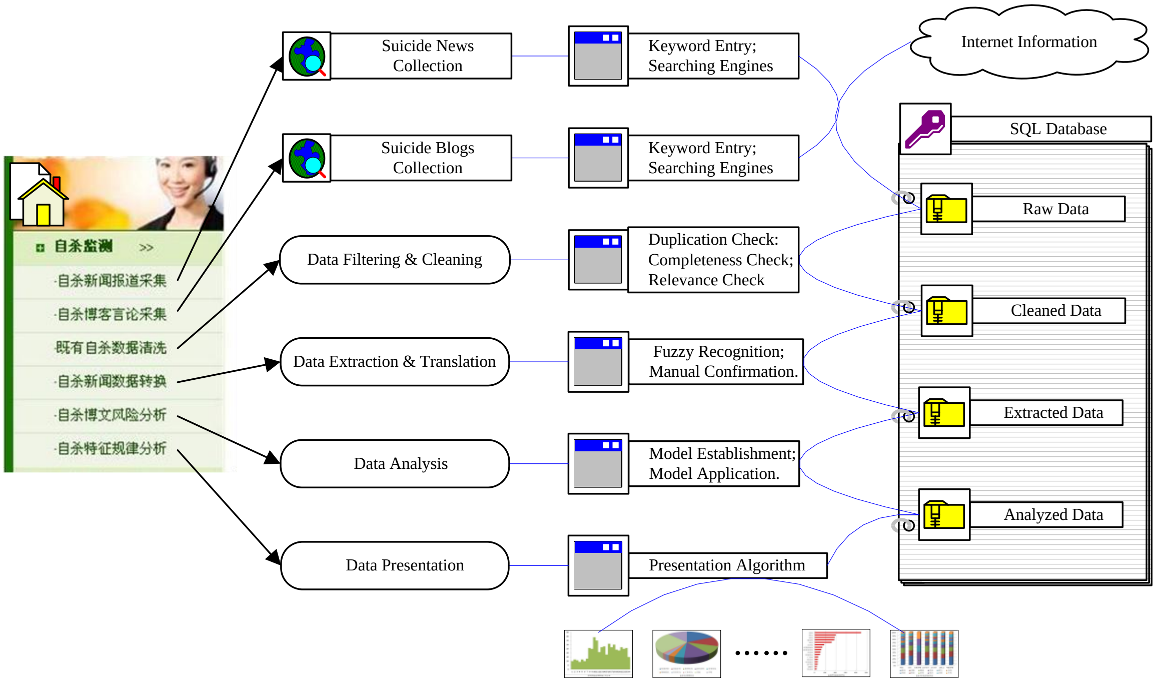

自杀监测 >>
·自杀新闻报道采集
·自杀博客言论采集
·既有自杀数据清洗
·自杀新闻数据转换
·自杀博文风险分析
·自杀特征规律分析
Suicide News Collection
Suicide Blogs Collection
Data Filtering & Cleaning
Data Extraction & Translation
Data Analysis
Data Presentation
Keyword Entry; Searching Engines
Keyword Entry; Searching Engines
Duplication Check: Completeness Check; Relevance Check
Fuzzy Recognition; Manual Confirmation.
Model Establishment; Model Application.
Presentation Algorithm
Internet Information
SQL Database
Raw Data
Cleaned Data
Extracted Data
Analyzed Data
……

Fig. 1 Overall architecture and data flow of SIMS

a skeletal SIMS; (d) evaluation and concept refinement through actual application of the skeletal SIMS; (e) implementation of the refined requirements; (f) cycles of steps (d) and (e) [16].

B. *Data Resource and Collection*

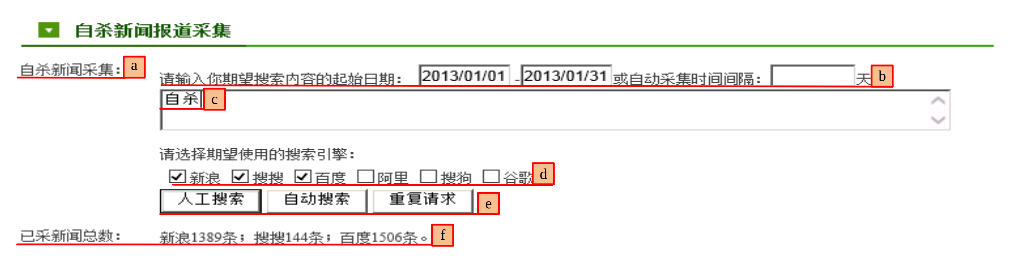
SIMS uses existing Chinese web-search engines for gathering primary suicide-related source data from the Internet. It captures, via its built-in window, user entered keywords (such as suicide, want to die, too boring to live etc.) and automatically inputs these keywords into the selected most popular Chinese search engines (Sina, Soso, Baidu, Sogou, Alibaba and Google) and triggers them to perform a systematic search for webpages containing the keywords via a short C# program using “HttpWebResponse” and “HttpWebRequest” as the key objects. All the web-pages retrieved by these search engines are traced and downloaded into a SQL database automatically for later use.

C. *Data Filtering and Cleaning*

The primary data gathered into the SQL database through the above mentioned process and program contain lots of incomplete, duplicated and irrelevant contents and thus need proper screening and cleaning. SIMS completeness check comes up with missing contents and comprises two steps: a) automatic screening for all very short webpages (with fewer than 100Chinese characters) gathered in the database; b) manual browsing of the original source of all the very short webpages by clicking the URL (Uniform Resource Locator) provided by SIMS and making necessary corrections if applicable. SIMS duplication check identifies and excludes duplicated contents according to the following steps and rules: a) if two webpages retrieved by different search engines or by a same engine were traced to a same original URL, only one of them is kept in the database; b)for the remaining webpages after a), if two of them have identical titles and less than difference in number of Chinese characters (after html labels have being excluded), only the one with earlier publication date remains in the database; c) for the remaining webpages after b), if two of them share over 70% of similar sentences, the one with later publication date is removed from the database. Here, “similar sentences” define sentences that share over 86% identical serial 2-grams (2 consecutive Chinese characters). SIMS relevance check eliminates irrelevant webpages searched using models developed trough: a) identification of keywords that helps in distinguishing relevant vs. irrelevant suicide webpages by 3 experienced researchers through manual reading of sample suicide-related webpages; b) supervised learning [17] using the keywords identified in the previous step, manually classified webpages samples as training sets and Perceptron algorithms similar to that described by Stormo et al [18].

D. *Data Extraction and Translation*

All the data cleaned above are in text format. SIMS data extraction and translation generates structured data from these unstructured texts so as to enable quantitative analysis. Main items of structured data SIMS currently produces include: a) place, date and time of webpage-reported suicide events; b) reasons, methods and results of the suicides reported; c) age and gender of the suicide victims; d) publication date, time and websites of blogs with high suicide ideations; e) causes to and degree of suicide ideations. Extraction of all these data, except for degree of suicide ideation, consists of “automatic fuzzy recognition” plus “manual confirmation”. Taking the



自杀新闻采集：a

请输入你期望搜索内容的起始日期：2013/01/01 - 2013/01/31 或自动采集时间间隔： 天 b

自杀 c

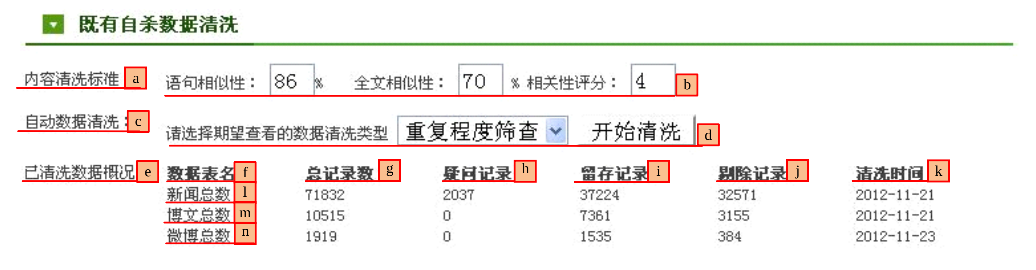
请选择期望使用的搜索引擎：

☑ 新浪 ☑ 搜搜 ☑ 百度 □ 阿里 □ 搜狗 □ 谷歌 d

已采新闻总数： 新浪1389条；搜搜144条；百度1506条。f

Ⅰ Suicide news collection

a: News collection; b: please input the start and end dates (for manual data collection) or interval (for automatic data collection); c: keywords (e.g. suicide) to be searched; d: alternative search engines; e: manual search/ automatic search/repeated request button; f: number of collected news from different engines



内容清洗标准 a 语句相似性：86 % 全文相似性：70 % 相关性评分：4 b

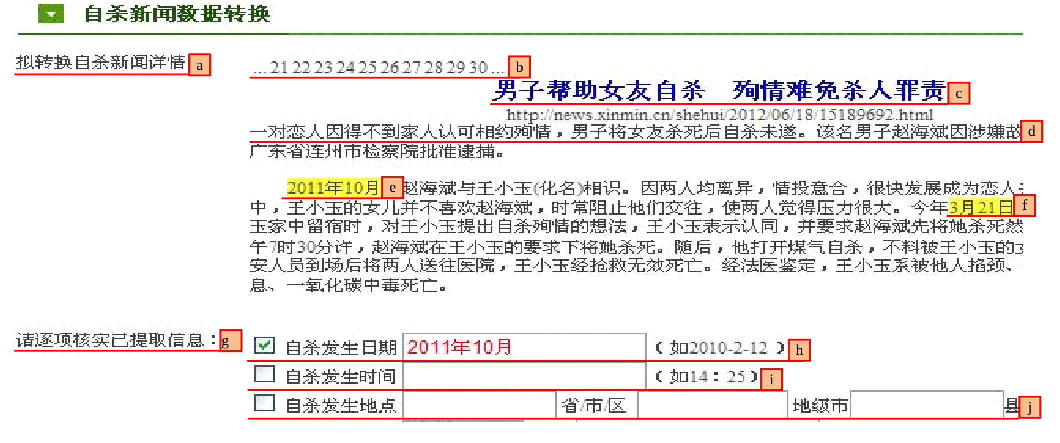
自动数据清洗：c 请选择期望查看的数据清洗类型 重复程度筛查 d

已清洗数据概况 e

数据表名 f	总记录数 g	疑问记录 h	留存记录 i	剔除记录 j	清洗时间 k
新闻总数 l	71832	2037	37224	32571	2012-11-21
博文总数 m	10515	0	7361	3155	2012-11-21
微博总数 n	1919	0	1535	384	2012-11-23

Ⅱ Data filtering and cleaning

a: criteria for data cleaning; b: sentence similarity (86%), full content similarity (70%), relevance score (4); c: automatic data cleaning; d: please choose type of data cleaning (completeness check) and start cleaning button; e: summary of cleaned data; f. data table name; g: total records; h. questionable records; i: remained records; j: excluded records; k: date of cleaning; l: number of news; m: number of blogs; n: number of microblogs



拟转换自杀新闻详情 a

... 21 22 23 24 25 26 27 28 29 30 ... b

男子帮助女友自杀　殉情难免杀人罪责 c

http://news.xinmin.cn/shehui/2012/06/18/15189692.html

一对恋人因得不到家人认可相约殉情，男子将女友杀死后自杀未遂。该名男子赵海斌因涉嫌故 d
广东省连州市检察院批准逮捕。

2011年10月 e 赵海斌与王小玉(化名)相识。因两人均离异，情投意合，很快发展成为恋人；
中，王小玉的女儿并不喜欢赵海斌，时常阻止他们交往，使两人觉得压力很大。今年3月21日 f
玉家中留宿时，对王小玉提出自杀殉情的想法，王小玉表示认同，并要求赵海斌先将她杀死然
午7时30分许，赵海斌在王小玉的要求下将她杀死。随后，他打开煤气自杀，不料被王小玉的
安人员到场后将两人送往医院，王小玉经抢救无效死亡。经法医鉴定，王小玉系被他人掐颈、
息、一氧化碳中毒死亡。

请逐项核实已提取信息：g

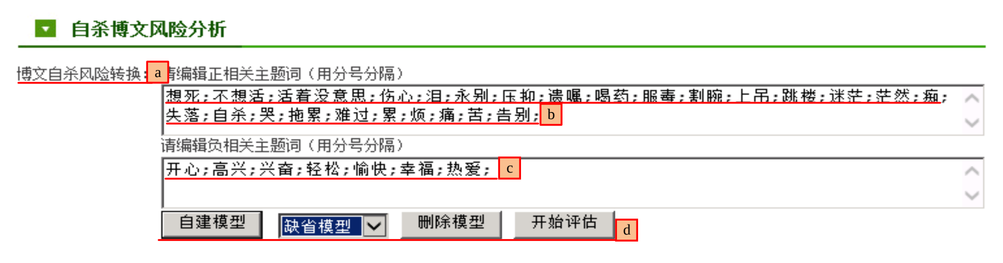
☑ 自杀发生日期 2011年10月 （如2010-2-12） h

□ 自杀发生时间 （如14：25） i

□ 自杀发生地点 省/市/区 地级市 县 j

Ⅲ Data transformation

a: detailed news reports; b: page number; c: title of news; d: sample sentence of news; e & f: highlighted text (of potential dates of suicide) g: please check extracted data item by item; h: suicide date (2011-10); i: suicide time; j: suicide place-province/city/county



博文自杀风险转换：a 请编辑正相关主题词（用分号分隔）

想死；不想活；活着没意思；伤心；泪；永别；压抑；遗嘱；喝药；服毒；割腕；上吊；跳楼；迷茫；茫然；痴；失落；自杀；哭；拖累；难过；累；烦；痛；苦；告别； b

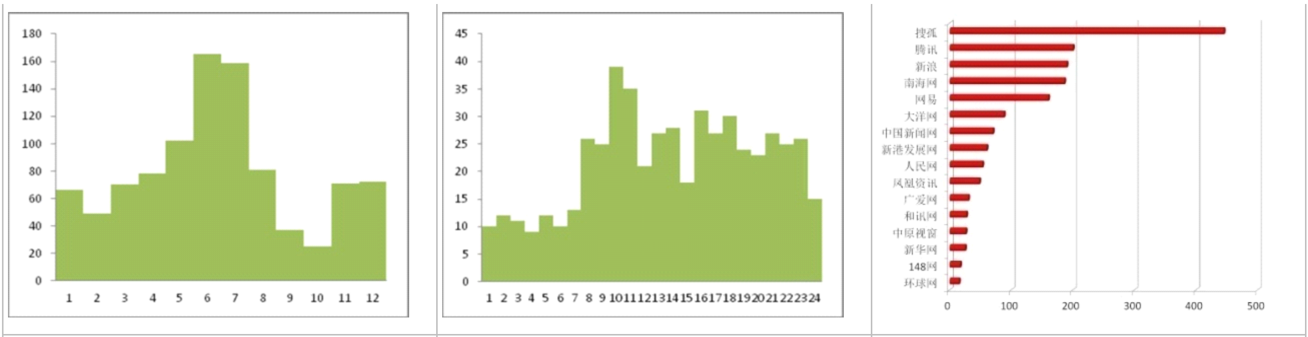
请编辑负相关主题词（用分号分隔）

开心；高兴；兴奋；轻松；愉快；幸福；热爱； c

Ⅳ Blog suicide ideation estimation

a: blog suicide ideation estimation; b: positively related terms to be included in estimation model; c: negatively related terms to be included in estimation model; d: build model, alternative models to be used (default model), delete model, start estimation

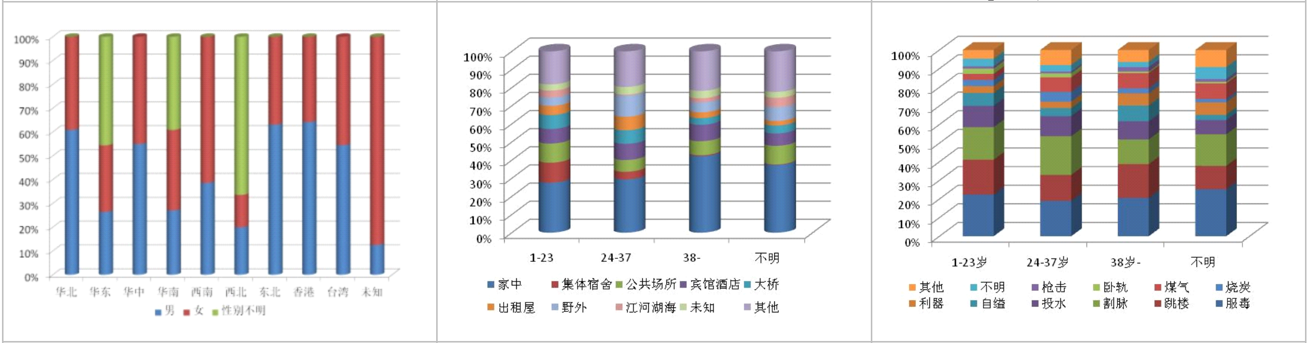
Fig. 2 Interfaces and general features of main SIMS functions



i. Suicide events by months (Y- axis: number of suicide events; X-axis: month)

ii. Suicide events by hours (Y-axis: number of suicide events; X-axis: hour)

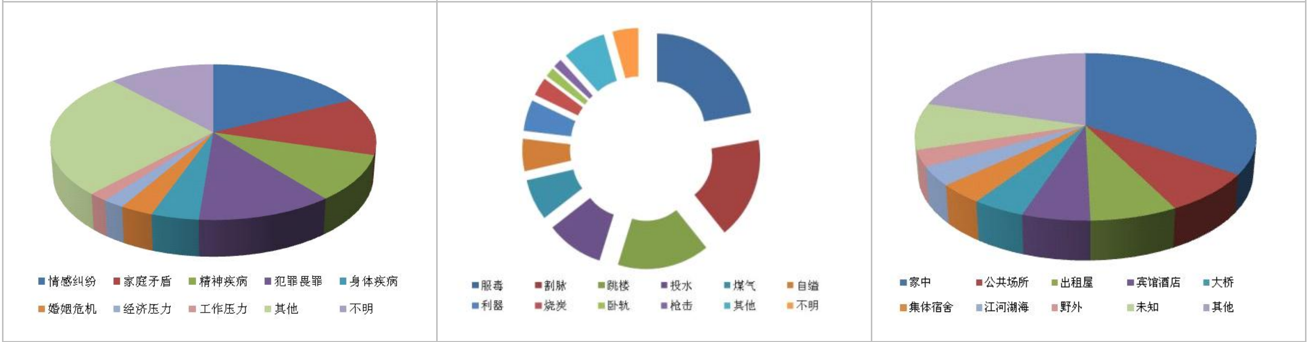
iii. Suicide events by sources (Y-axis: name of websites; X-axis: number of suicide reports)



iv. Suicide events by genders and regions (Y-axis: percentage; X-axis: name of geographic divisions e.g. north China, east China, etc.)
■ male ■ female ■ unknown

v. Suicide events by ages and places (Y-axis: percentage, X-axis: age)
■ home ■ dormitory ■ public place ■ hotel
■ bridge ■ rented house ■ wild field ■ river/ lake
■ others ■ unknown

vi. Suicide events by ages and methods (Y-axis: percentage, X-axis: age)
■ unknown ■ other ■ shoot ■ lie on rail
■ gas ■ burn charcoal ■ knife ■ hang
■ drown ■ cut vein ■ jump ■ poison



vii. Suicide events by reasons
■ emotional conflicts ■ family conflicts ■ mental diseases
■ criminal fear ■ physical diseases ■ marriage crisis
■ economic pressure ■ work pressure ■ others
■ unknown

viii. Suicide events by methods
■ poison ■ cut vein ■ jumper ■ drown
■ gas ■ hanger ■ knife ■ burn charcoal
■ lie on rail ■ shoot ■ other ■ unknown

ix. Suicide events by places
■ home ■ public place ■ rented house ■ hotel
■ bridge ■ dormitory ■ river / lake ■ wild field
■ others ■ unknown

Fig. 3 Sample presentations of system outputs.

example of extracting "date" of suicide from a given piece of webpage news stored in the database, SIMS automatically identifies and highlights all the fragments of texts that match preset "regular expressions" [19] within the piece of news under concern and picks the most likely or the first fragment and puts it into the entry "textbox" for manual checking and revision by relevant researcher or user. With regard to estimation of suicide ideation, it adopts a model developed via similar methodology and steps as that of relevance check described earlier.

E. Data Analysis and Presentation

At present, SIMS data analysis consists of mainly frequencies and composition calculations. All these calculations are programmed using structured query language (SQL) [20] developed by Microsoft. While presentation of these analysis results take two major formats, i.e., tables and diagrams. Tabular presentation employs mainly "ASP.NET" built-in controls (e.g., "GridView", "ListView"); while diagrams are drawn by self-designed programs based on the object entitled "System. Drawing" imbedded in Visual Studio 2008 and following the data visualization styles used by Microsoft Excel.

3. Results

A. Overall Architecture and Functions

The overall architecture of SIMS is shown in Fig. 1.It was designed to facilitate suicide monitoring via internet information (text-format news and blogs at present). Potential users of SIMS may include national, provincial, municipal and other regional Disease Control and Prevention Centers (CDC), regional mental health centers and institutions for suicide researches. SIMS comprises five main functions including suicide news collection, suicide blogs collection, data filtering and cleaning, data extraction and translation, blog suicide ideation estimation and data analysis and presentation.

B. Function Specific Features

Fig. 2 shows general features of specific SIMS functions. Data collection provides a user-friendly interface for retrieving suicide-related news and blogs from Chinese webpages and downloading them into SIMS database, in which all what users need to do is only entering relevant keywords and start and end dates (for manual data collection) or time interval (for automatic data collection) and click relevant buttons. Data filtering and cleaning comprises completeness, duplication and relevance checks. For duplication and relevance checks, a user can either enter his or her preferred cut points of "full text similarity", "sentence similarity" and "relevance score" first and then select relevant type of data cleaning or select and click the controls directly if he or she prefers the default cutting points set by SIMS. Completeness check is a bit more cumbersome since it requires manually browsing the source webpages of too short (e.g., less than 100 Chinese characters) news or blogs retrieved by SIMS. Data extraction and translation is mainly a manual task yet SIMS facilitates this work greatly by locating and highlighting the text fragments that most likely contain the information the users are searching for. Similarly, blog suicide ideation estimation also enables default or user defined model application. By default model using, suicide ideation estimation is accomplished by simply clicking the "start estimation" button. With regard to suicide ideation estimation via user defined models, it proceeds in three steps: a) editing the default positive and negative terms of suicide and clicking the "start learning" button to build a user preferred estimation model; b) clicking the "save model" button to add the new model into the "DropDownList" of alternative model list; c) selecting a preferred model from the list and clicking the "risk estimation" button to generate suicide ideation scores for the blogs to be analyzed.

C. System Outputs

The outputs of SIMS include mainly frequencies and compositions of webpage reported suicide events and blogs with high suicidal ideations published in total and by different genders, ages, regions, causes, methods, years, months and hours etc. SIMS adopts a paired graph and table strategy in presenting these indicators. The graph visualizes the trend or composition of a specific indicator under concern; while the table allows more detailed examination and interpretation of the indicator. In order to allow observing findings from different perspectives, SIMS provides a set of graphs for user to choose (by selecting relevant options of the graph type "DropDownList") including histogram, pie chart and plotting chart. Fig. 3 gives sample graphs.

4. Discussion

A. Usefulness of SIMS

SIMS provides a novel yet practical means for monitoring and understanding suicide. This is especially true for nations or regions, like China, where lack adequate traditional suicide reporting systems and epidemiologic surveys. Traditional reports and surveys require huge resource input [21]; while digging suicide related information from news and blogs on the internet published by professional reporters or ordinary people is almost costless. Even in areas with well-developed suicide reporting systems, suicide webpages may still serve as an important complementary information source. Traditional reporting relying on death registry and clinical services include only the suicide events that have resulted in death or injuries needing clinical treatments [22], while omitting attempted suicides without severe injuries. Traditional surveys may reveal some suicide ideations, processes and experiences but have to be highly focused due to resource limits leaving various aspects uncovered, e.g., continuous changes in suicide ideation and related emotional states of the victims, indirect causes leading to suicide plans and practice, feelings and observations of bystanders [23].

The outputs SIMS produces can be helpful in various perspectives. The bar charts of source websites, for instance, suggest where to look for suicide-related data and which are the priority venues for disseminating prevention information and educating more appropriate writing of suicide-related news, blogs etc. Similarly, the histograms depicting the frequencies of webpage-reported suicide events by hours and months provide clues for timing future intentions against suicide; the composition graphics (pies, circles, lines etc.) of suicidal events and blogs by genders, ages, regions and suicide causes and methods yield implications for identifying high or low risk groups and regions and pinpointing suicide causes and methods; while the list of blogs rank-ordered by estimated suicidal ideations locates high risk individuals on the internet and hints potential web-based and individualized interventions for them.

Another point worth noting is that SIMS not only helps in understanding and stemming suicide epidemic, but also informs interventions against other diseases [24-28]. The overall architecture and the strategies and techniques for data retrieval, storage, cleaning, analysis and presentation used by SIMS can be easily adjusted to build similar systems for tackling a whole range of similar diseases or health problems like traffic accidents, injuries, food poising, environment pollution events etc.

B. Lessons from SIMS

Successful development and application of SIMS depends on innovative tackling of a few fundamental challenges. The first challenge concerns completeness of data retrieved. On one hand, fulfilling SIMS purposes requires finding relevant suicide information from the internet as much as possible. On the other, in addition to socio-cultural and psychological factors that contribute to selection biases to be discussed later, there are technical barriers to complete data retrieval. For example, numerous websites contain suicide information and they employ different programming languages, architectures and coding standards of Chinese characters etc. In coping with this complexity, SIMS opts to use existing most popular and thus most powerful Chinese search engines rather than developing a standalone search tool for its own. This not only saves efforts developing the system but also insures maximum retrieval of wanted data. Another barrier to data completeness originates from the fact that a lot of websites keep only most recent (generally last 1-2 months) data and that some website servers may be inaccessible for various reasons (e.g. maintenances) at the time when users are retrieving data. These explain why SIMS allows automatic periodical data retrieval and repeated data requests from failed URLs.

The second challenge refers to duplication of data retrieved. According to our preliminary application, this duplication took three forms i.e., duplications due to same documents been searched by different engines, duplication due to same documents being replicated on different websites with little or certain modifications and the former two account for over 95% of total duplications. Given these, SIMS adopts a three-step approach in response, i.e., finding and excluding duplicated news or blogs by means of identical source URLs, similar titles and similar full contents respectively. Here, both the techniques used and the sequence of technique using merit mentioning. In fact, the third step alone can solve all duplication problems. Yet calculating the full content similarity between two documents requires separating the whole contents of the documents into two sets of sentences and then comparing the sentences within the two sets one by one. This involves a huge volume of computing. While comparing the source URLs and titles of two documents is much easier. So placing URL and title similarity check before full content similarity check can greatly expedite duplication reduction.

The third challenge relates to relevance and suicide ideation estimation. As described in methodology section, SIMS utilizes learned models to distinguish relevant from irrelevant news reports about suicide events and estimate ideation of blogs. This turned out to be more effective than we had thought of. For news relevance models, the key lies in manually read adequate original news reports gathered via SIMS, classify both relevant and irrelevant ones (rather than relevant news reports only) into sub-categories, and identify a few indicative keywords or phrases for each category. Using so derived keywords and adequate training set (e.g., a few hundred manually labeled relevant and irrelevant news reports), it is not difficult to get sufficient models, e.g., over 70%sensitivity and specificity, in serving users' purposes. Similarly, development of log suicide ideation estimation models rely on properly defined positive plus negative terms and enough rated blogs by experienced researchers.

The fourth challenge comes from extracting structured data from free texts. This includes selecting proper items of data to be extracted and designing effective algorithms locating the items in unstructured texts. The former depends heavily on incorporation of expertise on suicide epidemiology and on availability of data items from webpages. With regard to automatically locating given items, it requires careful manual reading of relevant webpages and find out all potential formats of presentations of the items. Given the fact that structured data are used for quantitative analysis and thus should be as accurate as possible, SIMS adopts an "automatic fuzzy recognition" plus "manual confirmation" strategy. This is a balanced decision between accuracy and expedience. Manual confirmation excludes potential mis-extracted data by SIMS; while fuzzy recognition reduces workload for the confirmation.

C. Limitations of SIMS

The outputs of SIMS derive from webpage suicide reports or blogs. Unlike data from mandate suicide reporting system and well-designed epidemiology surveys with clearly defined items of information to be collected, these reports or blogs are published under free choices by all kinds of writers, making selection biases inevitable and users are fully cautioned about these biases when interpret SIMS outputs. The models SIMS uses in distinguishing relevant from irrelevant news reports and estimating blog suicide ideations often result from compromises between sensitivity and specificity. In other words, users should not expect a hundred percent accurate estimation in building their own models. Instead, they should aim for models sufficient to serve their purposes. They should also estimate the sensitivities and specificities of the models they are using and adjust the outputs accordingly if necessary. Perhaps, the most difficult limitation with SIMS relates to its requirement of substantial manual work in extracting structured data, though the system greatly facilitates this process. This is closely linked with limited accuracy of SIMS models and fuzzy recognition algorithms and has little room for improvement in the short run. Finally, SIMS is still a very primitive helping tool and a variety of functions or applications can be added. Incorporating links analysis, for example, may help users trace how suicide news and blogs are transmitted and how to better disseminate prevention messages via the internet; while terms analysis, may reveal patterns or hints on what the general public are most concerned about suicide events and what feelings, activities, things etc. are most commonly associated with suicide ideations.

5. Conclusion

SIMS provides a novel and practical means for monitoring and understanding suicide. It proposes useful aspects as well as tools for analyzing the features and trends of suicide using data derived from the internet as supplements to traditional suicide reporting and epidemiology surveys. Although SIMS was designed specifically for suicide, the overall architecture, strategies and techniques can be easily adapted to similar systems for other diseases or health problems.

Acknowledgment

This paper was co-funded by the Natural Science Foundation of China (grant number 81172201) and Anhui Provincial Fund for Elite Youth (grant number 2011SQRL060).

Penglai Chen and Jing Chai contributed equally to this manuscript.

References

[1] Fleischmann, Alexandra, and Shekhar Saxena. Suicide Prevention in the WHO Mental Health Gap Action Programme (mhGAP). The Journal of Crisis Intervention and Suicide Prevention 2013; 34(5): 295–296. doi:10.1027/0227-5910/a000214.

[2] Keshavan, Matcheri S, Sudhakar Shenoy, and Huijun Li. Suicide in Asian Countries. Asian Journal of Psychiatry2013 Oct; 6(5): 355. doi:10.1016/j.ajp.2013.08.063.

[3] Zhang, Jie, and Lin Lin. The Moderating Effects of Impulsivity on Chinese Rural Young Suicide. Journal of Clinical Psychology2013 Sep 3; doi:10.1002/jclp.22039.

[4] Simon, Melissa, E-Shien Chang, Ping Zeng, and Xinqi Dong. Prevalence of Suicidal Ideation, Attempts, and Completed Suicide Rate in Chinese Aging Populations. A Systematic Review: Archives of Gerontology and Geriatrics 2013 Dec;57(3): 250–256. doi:10.1016/j.archger.2013.05.006.

[5] Lu, Jin, Yize Xiao, Xiufeng Xu, Qingping Shi, and Yunjuan Yang. The Suicide Rates in the Yunnan Province, a Multi-ethnic Province in Southwestern China. International Journal of Psychiatry in Medicine 2013; 45(1): 83–96.

[6] Zhao, Jiu-Bo, Jing-Bo Zhao, Rong Xiao, Xue-Ling Yang, and Xiao-Yuan Zhang. [Suicide exposure and its modulatory effects on relations between life events and suicide risk in Chinese college students]. Nan fang yike da xuexuebao = Journal of Southern Medical University2013 Aug 20; 33(8): 1111–1116.

[7] Hagihara, Akihito, Takeru Abe, Megumi Omagari, Midori Motoi, and Yoshihiro Nabeshima. The Impact of Newspaper Reporting of Hydrogen Sulfide Suicide on Imitative Suicide Attempts in Japan. Social Psychiatry and Psychiatric Epidemiology2013 Jul 14; doi:10.1007/s00127-013-0741-8.

[8] Yang, Albert C, Shi-Jen Tsai, Norden E Huang, and Chung-Kang Peng. Association of Internet Search Trends with Suicide Death in Taipei City, Taiwan, 2004-2009. Journal of Affective Disorders 2011 Jul; 32(1–2): 179–184. doi:10.1016/j.jad.2011.01.019.

[9] Schwartz, Karen D, and Zana Marie Lutfiyya. In Pain Waiting to Die': Everyday Understandings of Suffering. Palliative & Supportive Care2012 Mar; 10(1): 27–36. doi: 10.1017/S1478951511000551.

[10] Lester, David.Linguistic Analysis of a Blog from a Murder-suicide. Psychological Reports 2010 Apr; 106(2):342.

[11] Ogburn, Kelin M, ErickMessias, and Peter F Buckley. New-age Patient Communications through Social Networks. General Hospital Psychiatry2011Apr; 33(2): 200.e1–3. doi:10.1016/j.genhosppsych.2010.08.006.

[12] Tatsis, Vasileios A, Christos Tjortjis, and Panagiotis Tzirakis. Evaluating Data Mining Algorithms Using Molecular Dynamics Trajectories. International Journal of Data Mining and Bioinformatics2013; 8, no 2: 169–187.

[13] Gurulingappa, Harsha, Luca Toldo, Abdul Mateen Rajput, Jan A Kors, Adel Taweel, and YorkiTayrouz. Automatic Detection of Adverse Events to Predict Drug Label Changes Using Text and Data Mining Techniques. Pharmacoepidemiology and Drug Safety2013 Aug 12; doi:10.1002/pds.3493.

[14] Patricia R, Samara E. Harms, Jeffrey M. Googling suicide: surfing for suicide information on the internet. J Clin Psychitary 2008; 69:878-888.

[15] Michael J.McCarthy. Internet monitoring of suicide risk in the population. Journal of Affective Disorders, 2010; 122: 277–279.

[16] Katase, Hiroshi, Manabu Kanazawa, Masanao Inokoshi, and Shunsuke Minakuchi. Face Simulation System for Complete Dentures by Applying Rapid Prototyping. The Journal of Prosthetic Dentistry 2013 Jun; 109(6): 353–360. doi: 10.1016/S0022-3913(13)60316-9.

[17] Cruz, Joseph A, and David S Wishart. Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Informatics 2006; 2: 59–77.

[18] Stormo, GD, TD Schneider, L Gold, and AEhrenfeucht. Use of the "Perceptron" algorithm to distinguish translational initiation sites in E. coli. Nucleic acids Research 1982 May 11; 10(9): 2997-3011.

[19] (DeLisle, Sylvain, Bernard Kim, Janaki Deepak, Tariq Siddiqui, AdiGundlapalli, Matthew Samore, and Leonard D'Avolio. Using the Electronic Medical Record to Identify Community-Acquired Pneumonia: Toward a Replicable Automated Strategy. PLoS ONE 2013 Aug; 8(8). doi:10.1371/journal.pone.0070944.)

[20] Moratilla, Juan M, Raul Alonso-Calvo, Gema Molina-Vaquero, Sergio Paraiso-Medina, David Perez-Rey, and Victor Maojo. A Data Model Based on Semantically Enhanced HL7 RIM for Sharing Patient Data of Breast Cancer Clinical Trials. Studies in Health Technology and Informatics 2013; 192: 971.

[21] Xing, Xiu-Ya, Fang-Biao Tao, Yu-Hui Wan, Chao Xing, Xiu-Yu Qi, Jia-Hu Hao, Pu-Yu Su, Hai-Feng Pan, and Lei Huang. Family Factors Associated with Suicide Attempts Among Chinese Adolescent Students: a National Cross-sectional Survey. The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine2010 June; 46(6): 592–599. doi:10.1016/j.jadohealth.2009.12.006.

[22] Ougrin, Dennis, Reetoo Banarsee, Valentine Dunn-Toroosian, and Azeem Majeed. Suicide Survey in a London Borough: Primary Care and Public Health Perspectives. Journal of Public Health 2011 Jan; 33(3): 385–391. doi:10.1093/pubmed/fdq094.

[23] Niederkrotenthaler, Thomas, Benedikt Till, Arno Herberth, Martin Voracek, Nestor D. Kapusta, Elmar Etzersdorfer, Markus Strauss, and Gernot Sonneck. The Gap Between Suicide Characteristics in the Print Media and in the Population. The European Journal of Public Health 2009 Jan; 19(4): 361–364. doi:10.1093/eurpub/ckp034.

[24] Moratilla, Juan M, Raul Alonso-Calvo, Gema Molina-Vaquero, Sergio Paraiso-Medina, David Perez-Rey, and Victor Maojo. A Data Model Based on Semantically Enhanced HL7 RIM for Sharing Patient Data of Breast Cancer Clinical Trials. Studies in Health Technology and Informatics 2013; 192: 971.

[25] Van Ballegooijen, Wouter, Heleen Riper, Britt Klein, David Daniel Ebert, Jeannet Kramer, Peter Meulenbeek, and Pim Cuijpers. An Internet-Based Guided Self-Help Intervention for Panic Symptoms: Randomized Controlled Trial. Journal of Medical Internet Research 2013 Jul 29; 15(7). doi:10.2196/jmir.2362.

[26] Yang, Albert C, Shi-Jen Tsai, Norden E Huang, and Chung-Kang Peng. Association of Internet Search Trends with Suicide Death in Taipei City, Taiwan, 2004-2009. Journal of Affective Disorders 2011 Jul; 132(1–2): 179–184. doi:10.1016/j.jad.2011.01.019.

[27] Ryhänen, Anne M., Sirkku Rankinen, Kirsi Tulus, Heikki Korvenranta, and Helena Leino-Kilpi. Internet based patient pathway as an educational tool for breast cancer patients. International Journal of Medical Informatics2012; 81(4): 270–278. doi:10.1016/j.ijmedinf.2012.01.010.

[28] Black, Erik, Jennifer Light, Nicole Paradise Black, and Lindsay Thompson. Online Social Network Use by Health Care Providers in a High Traffic Patient Care Environment. Journal of Medical Internet Research 2013 May 15; 15(5). doi:10.2196/jmir.2421.