

Intelligent Agent for Business Mail Server Administration for Terraqueous Regnant Business Entities Using Knowledge Discovery in Text

Venkatesh T¹ and Mahendra Varman R²

¹Department of Computer Science and Engineering, *SMK Institute of Tech*, Chennai, India

²Department of Computer Science and Engineering, Sri Sairam Engineering College, Chennai, India

Abstract

With the advent of technologies in NLP and Data Mining we can use its potential for a wide platform of application. In Business domain, the customer complaint and feedback is one walk where intense care and work and labour is needed. So, it's usually outsourced, which has its own advantage. We propose an idea to implement a better solution for the above problem using the stated technology.

Keywords: E mail, Classification, Service desk automation, Knowledge Discovery in Text, Data Mining, Text mining, Artificial Intelligence

1. Introduction

The growth of computer technology has its influence in all fields. All corporates have understood it's the only way to reach the El Dorado of profit, time, and quality. And Mining is one such technology which would help them in saving money, time, and labour. We here provide an intelligent agent to solve their mail management problem. We use the existing data mining tools and technologies on a mail server to build an intelligent agent for the betterment of the firm.

2. Type Style and Fonts

The basic way in which mail segregation is done is a lengthy process so the company usually outsources it to a BPO. Now there lies a risk of the BPO personnel not understanding the application of the domain problem and addresses it to the wrong person then it has to be reprocessed from the scratch. Usually may take a day or two and it also has possibility for the BPO to result into a business nemesis. But this agent would be a novel solution.

3. Ease of Use

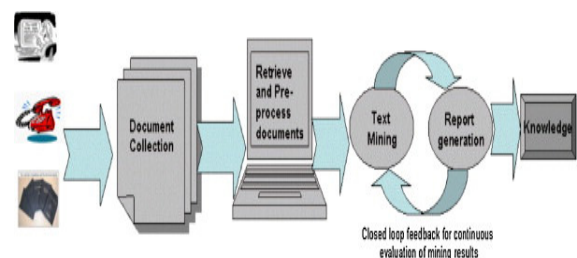
Text mining has been defined as "the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources". Text mining is similar to data mining, except that data mining tools are designed to handle structured data from databases or XML files, but text mining can work with unstructured or semi-structured data sets such as emails, full-text documents, HTML files, etc. As a result, text mining is a much better solution for companies, where large volumes of diverse types of information must be merged and managed.

Thus in this paper we use the potential for predictive mining and classify the data according to the user's needs, and not as an independent document. Please do not revise any of the current designations.

4. Prepare Your Paper Before Styling

The KDT technology can be used to classify the text document based on the genre by analysing the keyword distribution. The business domain of application would be, Exploiting this technology on the mails sent to a service desk.

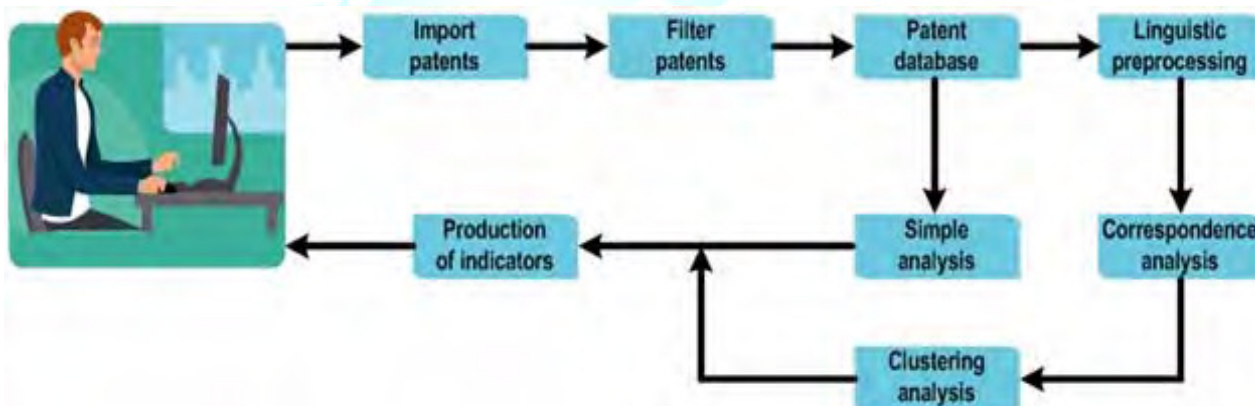
Taking the business domain is to be of servicing the User desk. The user service desk receives a lot of mails in a day and the service personnel reads them manually and forwards it to the respective desk. And a problem being



the Down time of the process and further if the mail is sent to a similar but different service dept then it has again to be 4waderd. To avoid this an automisation techneque has o be deviced. We use Data Mining(KDT) for this for this.

Purpose of text mining

1. To discover and use knowledge that is contained in a document collection as a whole, extracting essential information from document collections and from a variety of different sources.
2. Text mining lets executives ask questions of their text-based resources, quickly extract information and find answers they never imagined.



- Use a zero before decimal points: “0.25”, not “.25”. Use “cm³”, not “cc”. (*bullet list*)

B. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in Eq. 1, using a right tab stop. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. for quantities and variables, but not Greek

symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in

$$\alpha + \beta = \chi.$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “Eq. 1” or “Equation 1”, not “(1)”, especially at the beginning of a sentence: “Equation 1 is . . .”

A. Units

- Use either SI or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in Oersted. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter”, not “webers/m²”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.

C. Some Common Mistakes

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0 , and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.

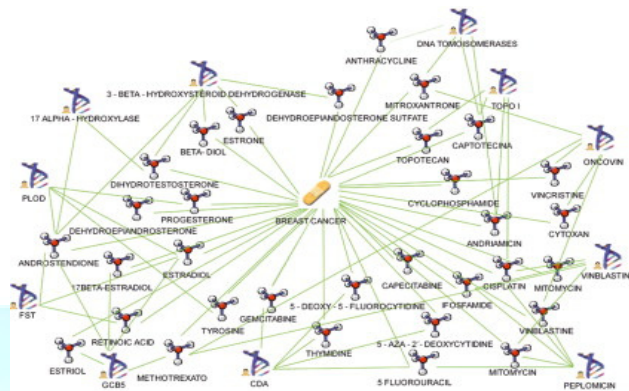
- In American English, commas, semi-/colons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively”.
- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.

Steps to Text Mining:

- "Preprocessing" the text to distill the documents into a structured format.
- Reducing the results into a more practical size.
- Mining the reduced data with traditional data mining techniques.

Text preprocessing transforms text into an information-rich, term-by-document matrix. This large grid indicates the frequency of every term within the document collection. During this stage, feature extraction is also used to locate specific bits of information, such as customer names, organizations and addresses.

Next, a mathematical technique called **singular value decomposition** (SVD) is used to replace the original term-by-document matrix with a much smaller matrix. As part of this process, unimportant words get discarded or ignored, and more important or highly relevant words are singled out. The new matrix can be used to place associated terms and documents into categories.



Next, **clustering, classification and predictive methods** are applied to the reduced data, using traditional data mining techniques. Conventional structured data sources can also be included in the analysis to enrich the discovery of underlying trends and patterns within the data.

Lastly, A snippet is coded in which the mined data is used to classified the mail in respective section, Further a mail id is mapped to the section which is used to forward the mail
 Conclusions/Applications: In this paper we put forward the concern of confidentiality and integrity of the mail and data which are a major deterrent for corporate enterprises. Thus it will help in fulfilling the customer relationship management needs. This being a prototype with minimal features can provide a base for better models of data maintenance and data manipulation using data mining.

References

- [1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (references)
- [2] H. Schutze. [Dimensions of meaning](#). In Proc. of the 1992 ACM/IEEE Conference on Supercomputing. IEEE Computer Society Press, Los Alamitos, CA, 1992, pp. 787–796
- [3] D. Lin. [Automatic retrieval and clustering of similar words](#). In Proc. of the 17th International Conference on Computational linguistics (COLING), Montreal, Canada, 1998, pp. 768–774.
- [4] D. Lin and P. Pantel. [Discovering word senses from text](#). In Proc. of the 8th International Conference on Knowledge Discovery and Data Mining (KDD), Edmonton, Canada, 2002, pp. 613–619.
- [5] D. Widdows and B. Dorow. [A graph model for unsupervised lexical acquisition](#). In Proc. of the 19th International Conference

- on Computational Linguistics (COLING), Taipei, Taiwan, 2002, pp. 1–7
- [6] J. Véronis. [Hyperlex: Lexical cartography for information retrieval](#). *Computer Speech and Language*, 18(3), 2004, pp. 223–252
- [7] ^ [a b c](#) Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "[From Data Mining to Knowledge Discovery in Databases](#)". Retrieved 17 December 2008.
- [8] "[Data Mining Curriculum](#)". [ACM SIGKDD](#). 2006-04-30. Retrieved 2011-10-28.
- [9] ^ Clifton, Christopher (2010). "[Encyclopædia Britannica: Definition of Data Mining](#)". Retrieved 2010-12-09.
- [10] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2009). "[The Elements of Statistical Learning: Data Mining, Inference, and Prediction](#)". Retrieved 2012-08-07.
- [11] Witten, Ian H.; Frank, Eibe; Hall, Mark A. (30 January 2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3 ed.). Elsevier. ISBN 978-0-12-374856-0.
- [12] Bouckaert, Remco R.; Frank, Eibe; Hall, Mark A.; Holmes, Geoffrey; Pfahringer, Bernhard; Reutemann, Peter; Witten, Ian H. (2010). "WEKA Experiences with a Java open-source project". *Journal of Machine Learning Research* **11**: 2533–2541. "the original title, "Practical machine learning", was changed [...] The term "data mining" was [added] primarily for marketing reasons."
- [13] ^ Mena, Jesús (2011). *Machine Learning Forensics for Law Enforcement, Security, and Intelligence*. Boca Raton, FL: CRC Press (Taylor & Francis Group). ISBN 9-781-4398-6069-4.
- [14] Piatetsky-Shapiro, Gregory; Parker, Gary (2011). "[Lesson: Data Mining and Knowledge Discovery: An Introduction](#)". *Introduction to Data Mining*. KD Nuggets. Retrieved 30 August 2012.
- [15] Kantardzic, Mehmed (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons. ISBN 0-471-22852-4. OCLC 50055336.
- [16] [Proceedings](#), International Conferences on Knowledge Discovery and Data Mining, ACM, New York.
- [17] [SIGKDD Explorations](#), ACM, New York.
- [18] Lukasz Kurgan and Petr Musilek (2006); [A survey of Knowledge Discovery and Data Mining process models](#). *The Knowledge Engineering Review*. Volume 21 Issue 1, March 2006, pp 1 - 24, Cambridge University Press, New York, NY, USA doi: 10.1017/S0269888906000737.
- [19] Günnemann, Stephan; Kremer, Hardy; Seidl, Thomas (2011). "An extension of the PMML standard to subspace clustering models". *Proceedings of the 2011 workshop on Predictive markup language modeling - PMML '11*. pp.48. doi:10.1145/2023598.2023605. ISBN 9781450308373. edit
- [20] Monk, Ellen; Wagner, Bret (2006). *Concepts in Enterprise Resource Planning, Second Edition*. Boston, MA: Thomson Course Technology. ISBN 0-619-21663-8. OCLC 224465825.