

# A Review on Material Recommendation System

Ms. K. Devika Rani Dhivya

Assistant Professor,  
Dept. of BCA & MSc.SS,  
Sri Krishna Arts and Science College,  
CBE, Tamil Nadu, India.  
Devika58@gmail.com

*Abstract*— Material Recommendation System discovers relevant feature in text documents and this evaluates the performance of student so that appropriate learning materials can be rendered to them. Text mining also referred to as text data mining, roughly equivalent to text analytics, refers to the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text analytics software can help by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed through different techniques. With an iterative approach, an organization can successfully use text analytics to gain understanding. Text analytics involves the application of techniques from areas such as information retrieval, natural language processing, and information extraction. These various stages of a text-analytics process can be combined together into a single workflow. Hence large industry dealing with enormous amount of text data are now easy for analyzes and interpretation. The main object is that the Easy Retrieval of relevant information's, Effective analyses and interpretation of text, faster data processing and accessing, Increases the degree of accuracy in retrieved data, Human intervention is reduced, large amount of processing time is minimized and Probability of missing any critical data is reduced.

## I. INTRODUCTION

Now a days students use a variety of strategies to make choices about what to buy, how to spend their leisure time, and even who to gain knowledge. Recommender systems automate some of these strategies with the goal of providing affordable, personal, and high-quality recommendations. Project and batch allocation, module coordinator upload project base paper on behalf of each and every student, and also allocate batches for all projects. Batches were created by coordinator by selecting number of student in batch and student ID's. In upload process the base paper will be checked for duplication with previous batches in title level as well as in content level. The content level checking is done by stripping down the pdf contents to text contents. If the base paper is not duplicated server accepts the upload and updates the student record. Text mining for assessment, Teacher prepares questions and answers for student assessment. Text mining process is done by natural language processing and word net tools. Pos tagger is implemented to extract the important keywords in the answer given by staff before assessment is done. The extracted Keywords are categorized mandatory keywords, subordinate keywords, and technical keywords. WordNet tool is used to give the related synonyms to literal word in the subordinate terms. Now Teachers can feed the servers with the eligible terms in the categories to be present for student evaluation.

Project Review and Student Assessment, Student login with his credentials and then uploads the review materials in server. Reviewer gives the review marks for each student based on performance. Here we allotted three reviews, and give marks for student based on review performance. Student can write the assessment test and can submit the answers to the server. Student answers are evaluated later in server by extracting keywords using NLP technique and WordNet tool. The Machine will evaluate the answers by comparing it with the categorized terms given by the teachers. Depends upon the student answer they will give marks and prepare performance report. Review performance and assessment score are aggregated to find the overall performance. Material Recommendation and Interactive Student learning, Teachers prepare the material for each subject and also give tags (good, best) for student material recommendation. Here we upload the materials like video, text, pdf. Video transcoding is applied while video materials are uploaded for below average students. After finishing the assessment test, in student portal they get the materials based on overall performance calculated by server. If they have doubt while watching video content, students can interactively raise questions by simply clicking on the video frame. The video frames are previous indexed so that appropriate Meta information's can be extracted for each frame. The student's questions and Meta information from the current frame are send to server and can be reviewed by the staff. Once the staff login they will be

notified with the questions and then staffs can reply to the question. The Student can now be able to view the answers given by the staffs.[1][2][8]

## II. EXISTING SYSTEM

The System learns term features within only relevant document and unlabeled documents, it used two term-based models. In the first stage, it utilized a Rocchio classifier to extract a set of reliable irrelevant documents from the unlabeled set. In the second stage, it built a SVM classifier to classify text documents. A two-stage model was also proposed in which proved that the integration of thorough analysis (a term-based model) and pattern taxonomy mining is the best way to design a two-stage model for information filtering systems.

### A. Pattern Taxonomy Mining

There are two challenging issues in using pattern mining techniques for finding relevance features in both relevant and irrelevant documents. The first is the low-support problem. Given a topic, long patterns are usually more specific for the topic, but they usually appear in documents with low support or frequency. If the minimum support is decreased, a lot of noisy patterns can be discovered. The second issue is the misinterpretation problem, which means the measures (e.g., “support” and “confidence”) used in pattern mining turn out to be not suitable in using patterns for solving problems. For example, a highly frequent pattern (normally a short pattern) may be a general pattern since it can be frequently used in both relevant and irrelevant documents. Hence, the difficult problem is how to use discovered patterns to accurately weight useful features. There are several existing methods for solving the two challenging issues in text mining. Pattern taxonomy mining (PTM) models have been proposed, in which, mining closed sequential patterns in text paragraphs and deploying them over a term space to weight useful features. Concept-based model (CBM) has also been proposed to discover concepts by using natural language processing (NLP) techniques. It proposed verb-argument structures to find concepts in sentences. These pattern (or concepts) based approaches have shown an important improvement in the effectiveness. However, fewer significant improvements are made compared with the best term-based method because how to effectively integrate patterns in both relevant and irrelevant documents is still an open problem.[1][2]

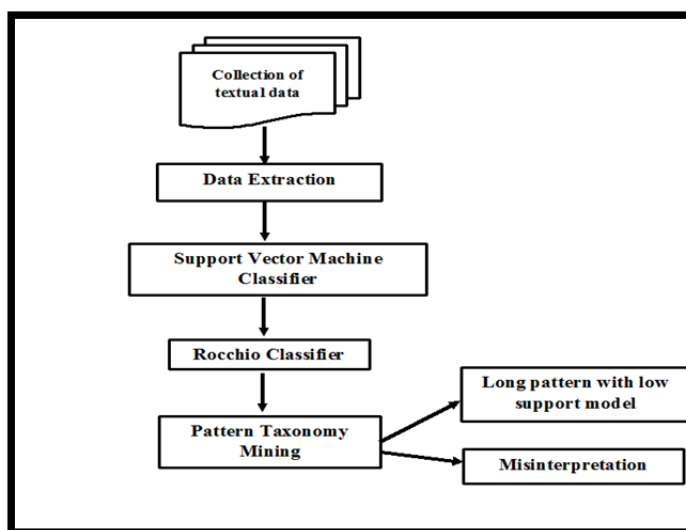


Fig .1 Process Flowchart

Data extraction is the act or process of retrieving data out of (usually unstructured or poorly structured) data sources for further data processing or data storage (data migration). The import into the intermediate extracting system is thus usually followed by data transformation and possibly the addition of metadata prior to export to another stage in the data workflow. Typical unstructured data sources include web pages, emails, documents, PDFs, scanned text, mainframe reports, spool files, classifieds, etc.

The act of adding structure to unstructured data takes a number of forms

- Using text pattern matching such as regular expressions to identify small or large-scale structure e.g. records in a report and their associated data from headers and footers.
- Using a table-based approach to identify common sections within a limited domain e.g. in emailed resumes, identifying skills, previous work experience, qualifications etc. using a standard set of commonly used headings (these would differ from language to language), e.g. Education might be found under Education/Qualification/Courses.

- Using text analytics to attempt to understand the text and link it to other information.

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

The static business models and static data models are used to model inherently dynamic business processes, particularly at the point of interaction. For example, virtually every customer relationship management system we have come across has a manual classification scheme (or taxonomy) that is meant to be used by the service agent to classify the nature of the student interaction. This approach has two major flaws. First, as soon as the classification scheme is published, it is out of date, because interactions with your customers are unpredictable and continually changing. Second, even if the classification scheme was representative of your customer interactions, it is unreasonable to expect any number of service agents to classify their interactions with their customers in a consistent way and with high quality. This very often makes such classification data completely useless, or, more dangerously, misleading. This issue is true throughout the business ecosystem where unstructured information exists.

It began when a potential faculty was considering a large unstructured data documents. Like most institutions, they had a huge collection of documents describing student interactions.

They wanted to automatically classify these documents to route them to the correct examination process. They questioned whether or not this was even feasible, and if so, how expensive would it be. Rather than invite all the vendors in this space to present proposals, they wanted to understand how effective each technical approach was on their data. A sample of 5,000 documents that had been scanned and converted to text and divided them manually into 50 categories of around 100 documents each. They then invited seven of the leading data mining vendors with products in this space to spend one week with the data using whatever tools and techniques they wished to model these 50 categories. When they were done, they would be asked to classify another unseen set of 25,000 documents. The different vendors' products would be compared based on speed, accuracy of classification, and ease of use during training. The results would be shared with all concerned.

#### ➤ *Disadvantages*

- No data accuracy.
- Low Data Retrieval.
- Classifying unstructured text is problematic.
- Pattern matching text is inaccurate.
- Amount of time spend is enormous.
- Mismatching of relevant data.

### III. PROPOSED SYSTEM

The material recommendation system based on student performance which will be predicted from the project work documents and the online examination results. The online tests are conducted for each individual subjects and the questions will be allocated for the same by the respective staff in their portal. The staffs will feed the evaluation knowledge to the system for student examination score generation. The aggregate of project work and online test decides the student's overall performance in three levels which is used to categorize them. The study materials will be rendered to the particular student depending upon the performance category he belongs. Videos can also be played in the student portal for effective learning which we innovate to interactive learning by simple click events on the doubtful content of the video.[3][4]

Natural language processing, often abbreviated as NLP, refers to the ability of a computer to understand human speech as it is spoken. NLP is a key component of artificial intelligence (AI) and relies on machine learning, a specific type of AI that analyzes and makes use of patterns in data to improve a program's understanding of speech.

- **Sentence segmentation, part-of-speech tagging, and parsing:** Natural language processing can be used to analyse parts of a sentence to better understand the grammatical construction of the sentence.
- **Machine translation:** Natural language processing is increasingly being used for machine translation programs, in which one human language is automatically translated into another human language.
- **Automatic summarization:** Natural language processing can be used to produce a readable summary from a large chunk of text. For example, one might use automatic summarization to produce a short summary of a dense academic article.

String searching algorithms, sometimes called string matching algorithms, are an important class of string algorithms that try to find a place where one or several strings (also called patterns) are found within a larger string or text. Video Transcoding, also called video encoding, is the process that converts a video file from one format to another, to make videos viewable across different platforms and devices.

Teacher prepares questions and answers for student assessment. Text mining process is done by natural language processing and word net tools. Pos tagger is implemented to extract the important keywords in the answer given by staff before assessment is done. The extracted Keywords are categorized mandatory keywords, subordinate keywords, and technical keywords. Wordnet tool is used to give the related synonyms to literal word in the subordinate terms. WordNet is a lexical database for the English language. It groups English words into sets of synonyms called synsets, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members. WordNet can thus be seen as a combination of dictionary and thesaurus. While it is accessible to human users via a web browser, its primary use is in automatic text analysis and artificial intelligence applications. The database and software tools have been released under a BSD style license and are freely available for download from the WordNet website. Both the lexicographic data (lexicographer files) and the compiler (called grind) for producing the distributed database are available.[7]

Teachers prepare the material for each subject and also give tags (good, best) for student material recommendation. Here we upload the materials like video, text, pdf. Video transcoding is applied while video materials are uploaded for below average students. After finishing the assessment test, in student portal they get the materials based on overall performance calculated by server. If they have doubt while watching video content, students can interactively raise questions by simply clicking on the video frame. The video frames are previous indexed so that appropriate Meta information's can be extracted for each frame. The student's questions and Meta information from the current frame are send to server and can be reviewed by the staff. Once the staff login they will be notified with the questions and then staffs can reply to the question. The Student can now be able to view the answers given by the staffs.

Now Teachers can feed the servers with the eligible terms in the categories to be present for student evaluation. The student's doubts are cleared by the faculty when the student stops the video, the video transcoding technique is implemented here to convert the video format into another format accessible by the reviewer. Thus with material recommendation system, the students capability is judged and relevant study materials are easily provided. This system helps to work accurately and efficiently with large amount of textual data. Thus material recommendation system based on student performance which will be predicted from the project work documents and the online examination results. The online tests are conducted for each individual subjects and the questions will be allocated for the same by the respective staff in their portal. The staffs will feed the evaluation knowledge to the system for student examination score generation. The aggregate of project work and online test decides the student's overall performance in three levels which is used to categorize them. The study materials will be rendered to the particular student depending upon the performance category he belongs. Videos can also be played in the student portal for effective learning which we innovate to interactive learning by simple click events on the doubtful content of the video.[6]

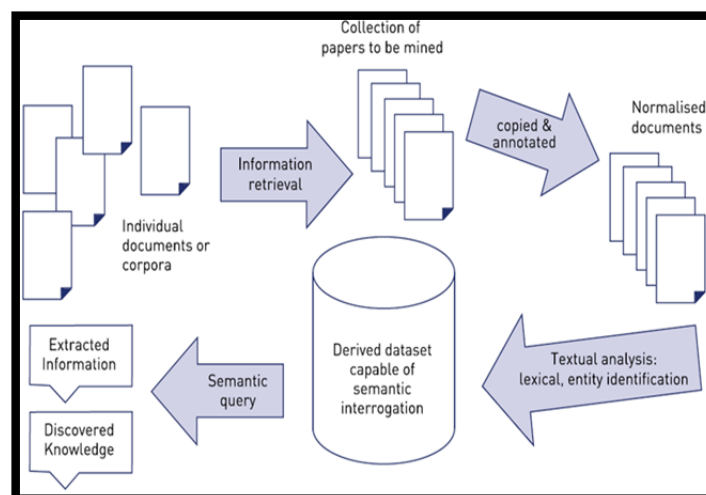


Fig 2 Discovering the Relevant Terms

➤ *Advantages*

- Easy Retrieval of relevant information's.

- Effective analyses and interpretation.
- Faster data processing and accessing.
- Increases the degree of accuracy in retrieved data
- Human intervention is reduced.
- Large amount of processing time is minimized.
- Probability of missing any critical data will be reduced.

#### IV. CONCLUSION AND FUTURE ENHANCEMENT

Hence the system developed a method to find and classify terms based on student's answer. Student test answer also classifies terms and evaluates answers based on teacher answer, and then teacher will provide material based on student performance. It presents a method to find and classify low-level features based on both their appearances in the higher-level patterns and their specificity. It also introduces a method to select irrelevant terms or documents for weighting features. In this way the project was developed and experimentally proved that the proposed specificity function is reasonable and the term classification can be effectively approximated by a feature clustering method. Thus project uses a feature clustering technique to automatically group terms into the three categories. The work is much more efficient and achieved the satisfactory performance as well. This work was thoroughly tested and the results prove that it is statistically significant. The system can be improved to match patterns containing more than a paragraph and produce higher quality information to the end users.

#### REFERENCES

- [1] Gary Miner "Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications" published 2012, 1<sup>st</sup> Edition.
- [2] Ghauth, K.I., Abdullah, N.A., Learning materials recommendation using good learners' ratings and content-based filtering, Educational Technology Research and Development, 2010;58(6):711-727.
- [3] García, E., Romero, C., Ventura, S., Castroa, C.d., A collaborative educational association rule mining tool, The Internet and Higher Education, 2011;14(2):77-88.
- [4] Symeonidis, P., Nanopoulos, A., Papadopoulos, Y. Collaborative recommender systems: Combining effectiveness and efficiency. Expert Systems with Applications, 2008;34(4):2996-3013.
- [5] Khribi, M.K., Jemni, M., Nasraoui, O., Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. Educational Technology and Society, 2009;12(4):30-42.
- [6] Kerkiri, T., Manitsaris, A., Mavridou, A., Reputation Metadata for Recommending Personalized e-Learning Resources, Second International Workshop on Semantic Media Adaptation and Personalization, 2007;110-115.
- [7] Fred Damerau, Nitin Indurkha, Tong Zhang, "Text Mining: Predictive Methods for Analyzing Unstructured information", published 2005.
- [8] Romero, C., Ventura, S., Zafra, A., Bra, P., Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. Computers and Education, 2009;53:828-840.
- [9] Luo, J., Dong, F., Cao, J., Song, A., A context-aware personalized resource recommendation for pervasive learning, Cluster Comput, 2010;13:213-239.
- [10] <http://www.onlineprogrammingbooks.com/textminingpos/>
- [11] <http://Text Mining: Predictive Methods for Analyzing Unstructured information -informations.com/>