Concept hierarchy extraction from textbook

Neha Pathak
M.E. Research Scholar, Department of C.S.
Institute of Engineering & Technology, D.A.V.V.
Indore ,India

Pragya Shukla
Professor: IET, DAVV, Indore
Computer Engg Dept.
pshukla@iet.dav.edu.in
pragyashukla_iet@yahoo.co.in

Nilima Karankar
Asst Professor: IET, DAVV, Indore
Computer Engg Dept.

***Abstract*— In modern day scenario, it is observed that the content available on Internet is in very large amount. Any single search will result in millions of results, thus the user faces the issues like which content to choose and which one to not. One of the classification in data mining is content hierarchy, which can be defined as the extraction of content in logical and related manner. The key focus of content hierarchy is extraction of knowledge from the data in order to derive the pattern. In order to perform the extraction, the quality and content is observed and should be present in relevant manner. In this work, the relevant content is first extracted and saved in the prerequisite, after it the next content which is highly relevant to the already saved data,stored in the form of hierarchy. The two concepts on which concept hierarchy is based, are local relatedness and global coherence. The application of the given system can be open educational system in which students and researchers will get the desired content. The current work lessen the burden of a reader by decreasing the entire book into the specified hierarchy and thus can be very much beneficial.**

Keywords—Data Mining;Semantic Annotation; Word Sense Disambiguation;

I. INTRODUCTION

Concept hierarchy extraction is based on the elicitation of important concept from wide verity of area. Concept is define as abstract idea of topic and an order of something formed by mentally combining all its characteristics. Hierarchy defines as a system in which words are ranked according to relative order. Extraction of concept gives us important idea about each topic of book. Concept hierarchy extraction in paper[1] follow learning to rank approach. Ranking is use to maintain order of perception in form of hierarchy. it basically work on two concept local applicability and global coherence. Local relatedness deal with wording closeness between chapters.

Global coherence gives relationship that concept in every chapter must be related to other chapter and no two chapter concept should be fully overlap. Concepts discuss between the chapters should be different. Concepts should be related to other concepts in same chapter with maintaining order of hierarchy.

Following are the Features of a good concept hierarchy system:-

1. Less redundancy: - Chapter content should be less similarity i.e. content have related to each other but not overlap concept.
2. Stability between chapters:- sub-chapter in chapters should be associated to each other.
3. Consistent learning order should follow sequential prerequisites i.e. Concept hierarchy should to follow structure of the knowledge.

II. LITERATURE REVIEW

Andrew M. Olney, Whitney L. Cade, and Claire Williams[1] defines meta analysis which gives parameter on the basis of learning gain and define relationship between conceptual b graph with predicate calculus. It defines approaches for extract concept from biological textbook following semnet formulation with some element. It define the concept map in form of relationship like is-a, has part.

Shutting Wang [2] Prerequisite Concept Maps Extraction for Automatic Assessment use Wikipedia to identify key concept extraction and content of article. it work on two phase title match and consin sim. Title match matches the title from Wikipedia to article. Consine sim matches concept vector between Wikipedia and article. This concept is works on top K candidate on match.

Judith b. Howard [3] proposed concept mapping as assessment tool a way of organizing idea about a particular topic. In this paper concept is divide in three part i.e. element, definition , point. As a tool for teaching and learning, concept mapping is used. This tool provides idea of assignment of every kind of relationship like 1point for each concept, 1point for each valid relationship, 3 point for each successive branching ,5point for each hierarchy, 10 point for each cross link,1point for each example this help in mapping concept and extracting important concept from book.

III. PROBLEM DOMAIN

In concept extraction we face multiple types of problem while extracting knowledge from text book and other domain. In broad domain like Wikipedia extraction of important concept from topic is very difficult task. Concept hierarchy 's biggest problem is related to global coherence where in paper [1] concept hierarchy extraction proposed solution related to global is np hard problem. To find relatedness and textual mapping between chapters in global coherence does not gives exact solution.

1. Global optimization computation for concept hierarchy is np hard problem and it gives approximation not exact answer.

2. Make use of search techniques for dictionary which is not guaranteed to find solution.

3.There is no exact representation of knowledge which make learning easy.

IV. SOLUTION DOMAIN

In our solution we use fuzzification to define relationship and dependency between chapters to build learning hierarchy.
Learning Hierarchies built for concept hierarchy, and hierarchies are designed by using expert knowledge. The relationship between the concept present and the learning hierarchy are not standard with absolute values indeed they are present in form of fuzzy relationships. These relationship thus generate the questions like is the establish relation of present perquisite correct or not ?. Hence the given solution is design in four phase:

1. Predefined learning hierarchy
2. Measure the variation of grades
3. Identify prerequisite between concepts
4. Exact relationship between concept
5. Propose to build the final learning hierarchy

A. Define an predefined learning hierarchy

To design any learning hierarchy by the experts of given field there is certain method and order, is defined by Gagne (1968) as mentioned below:

Set of concept dependency can be called as learning hierarchy. The relationship present is in ordered manner, as well as to have the effective instruction there should be some dependency between relation . The requirement is low level concepts must be mastered before the upper level concept. In the hierarchical model, the meeting point assigned are the concept .

In figure[1] defines relationship of concepts. A,B,C,D E,F,G defines as concept and there is dependency between concepts like B, C dependent on A and A is the prerequisite of B and C. To Understand concept of B,C we have to understand concept of A. T he lower level concepts as A and the upper level is B,C. Every high level concept is dependent on low level and the dependency hierarchy between chapters is given below:-

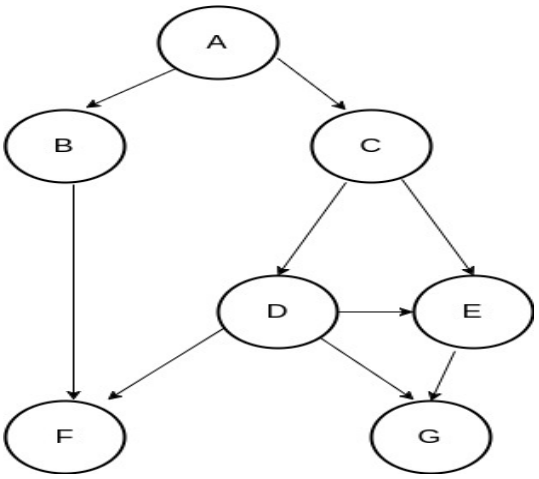

Figure1: Predefine learning hierarchy

In learning hierarchy prerequisite is define in form of i,j . which is represent in tabular form named as T. The calculation is define as:-

T_{ij} = 1 conveys the concept (i) essential and precondition of the concept (j).
T_{ij} = 0 conveys the concept (i) is not an essential and precondition of the concept (j).

Given below is the matrix representation T_{ij} of predefined learning hierarchy in Table 1 .

TABLE 1. MATRIX REPRESENTATION OF PREDEFINED LEARNING HIERARCHY

T_{ij}	A	B	C	D	E	F	G
A	0	1	1	0	0	0	0
B	0	0	0	0	0	1	0
C	0	0	0	1	1	0	0
D	0	0	0	0	1	1	1
E	0	0	0	0	0	0	1
F	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0

B. Measure the Changes in grades

In this stage the grades of each student are observed. The assessment are present in the form of matrix. The given figure[3] represent ten concept of the learning hierarchy of a course. If grades is represent as S_i and concept as "i" as:

In this the calculation of variation of grade is measured for having prerequisite relationship of predefined learning hierarchy.
The changes in grades is estimated by matrix:

- Relationship between (Si) and (i) Grades (Learner (Si), concept (i))
- Predefined learning hierarchy in matrix T_{ij}
- Changes in grades is define by as ΔGrades (i, j) Learner = [Grade (j) – Grade (i)] with T_{ij} = 1 i.e the concept (i) essential and precondition of the concept (j).
    - Grades measure between -20 ≤ Changes in grades(ΔGrades)≤ 20

Given below is the table which show Change in Grades (i, j) b as by expert system in TABLE 1 and TABLE 2:

The given figure represent ten concept of the learning hierarchy of a course define grades on concept on the basis of learning

TABLE 2. DEFINE VARIATION OF GRADES (ΔGRADES)

Grades	A	B	C	D	E	F	G
S_1	10	10	1	3	7	9	3
S_2	11	12	5	7	11	11	7
S_3	10	11	5	3	8	10	5
S_4	13	10	6	6	10	10	10
S_5	15	18	10	12	16	16	15
S_6	19	18	6	10	14	19	13
S_7	12	11	1	5	6	10	4
S_8	3	4	0	2	5	7	5
S_9	15	16	6	10	11	18	13
S_{10}	12	14	5	3	0	13	0

3. Identify prerequisite between concepts

The fuzzy logic is used to identify degree of each link to the membership function, and this membership function is used to map and identify prerequisite relationship between concepts.

Let P define as set of prerequisite relationships between concepts.

CPR can be define as correct order of prerequisite order between i and j .

CPR = $\{k , \mu_{cpr}(k)/ \quad K \in P\}$

The membership of CPR defines by function $\mu_{CPR}(k)$ and membership function show the relative degree of individual link k in fuzzy system.

RPR define as wrong prerequisite of relationship a fuzzy subset of links define as not correct essential and precondition order between concept (i)and concept (j)

RPR = $\{k , \mu_{RPR}(k)/ \quad K \in P\}$

$\mu_{RPR}(k)$ is the membership function of RPR, the

Result of this function show the relative degree of

Individual link (k) in the fuzzy set .

The two membership functions of fuzzy sets $\mu_{CPR}(k)$ and $\mu_{RPR}(k)$ is define as variation of grades in prerequisite relationships of predefined concept hierarchy .

C. Exact relationship between concept

For Exact relationship between the concept. The two $\mu_{CPR}(k)$ and μRPR (k) are based on following table:

Rule	Prerequisite relationships (k)
$S1 \leq \Delta Grades \leq S2$ $\{S1 < 0, S2 > 0\}$	$k \in CPR$
$S2 \leq \Delta Grades \leq S3$ $\{S3 > S2\}$	$k \in RPR$

On the basis of prerequisite rule $\mu_{CPR}(k)$ and $\mu_{RPR}(k)$ are define as in form of relationship between concept:-

$$\mu_{CPR}(k) = \begin{cases} 0 & \text{if } \Delta Grades < S1 \\ \frac{-1}{S1}\Delta Grades + 1 & \text{if } S1 \le \Delta Grades \le 0 \\ \frac{-1}{S2}\Delta Grades + 1 & \text{if } 0 < \Delta grades \le S2 \\ 0 & \text{if } \Delta Grades > S2 \end{cases}$$

$$\mu_{RPR}(k) = \begin{cases} 0 & \text{if } \Delta Notes < 0 \\ \frac{1}{S2}\Delta Notes & \text{if } 0 \le \Delta Notes \le S2 \\ \frac{-(\Delta Notes + S3)}{S3 - S2} & \text{if } S2 < \Delta Notes \le S3 \\ 0 & \text{if } \Delta Notes > S3 \end{cases}$$

We define the threshold which is S1, S2 and S3 are defined in association with experts system. Based on analysis.
Values are define as follows:

S1 = change in -5 grades

S2 = change in 5 grades

S3 = change in 10 grades

Function $\mu_{CPR}(k)$ and $\mu_{RPR}(k)$ with grades define as expert system evolution:

$$\mu_{CPR}(k) = \begin{cases} 0 & \text{if } \Delta Grades < -5 \\ \frac{1}{5}\Delta Grades + 1 & \text{if } -5 \le \Delta Grades \le 0 \\ \frac{-1}{5}\Delta Grades + 1 & \text{if } 0 < \Delta grades \le 5 \\ 0 & \text{if } \Delta Grades > 5 \end{cases}$$

$$\mu_{RPR}(k) = \begin{cases} 0 & \text{if } \Delta Grades < 0 \\ \frac{1}{5}\Delta Grades & \text{if } 0 \le \Delta Grades \le 5 \\ \frac{-1}{5}\Delta Grades + 2 & \text{if } 5 < \Delta grades \le 10 \\ 0 & \text{if } \Delta Grades > 10 \end{cases}$$

Experiments and Results

We conduct experimental analysis to elicit concept hierarchy in different area. Our experiment is based on whether the proposed fuzzyfication technique effective for extract important concept in each chapter.

We conduct set of experiment foe making concept hierarchy. The experiment on book use one books as experiment data and chapter level analysis .we finally gives study on concept hierarchy extraction from the java book.

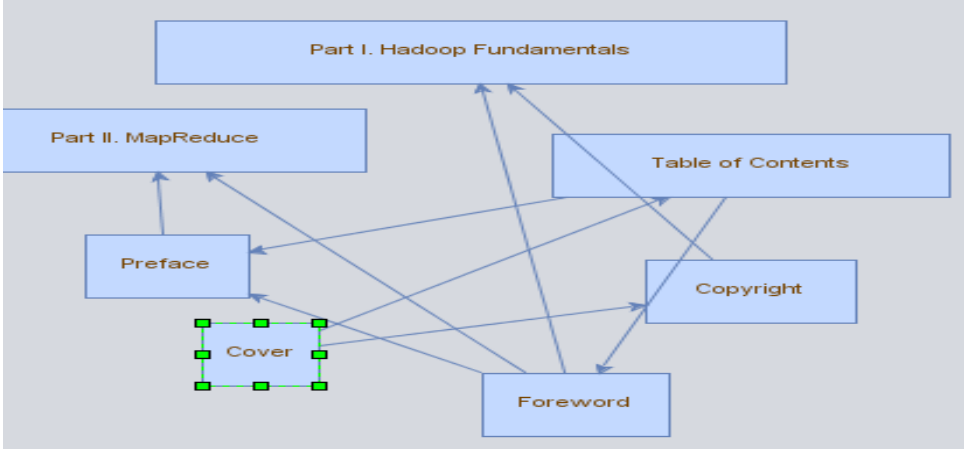

Figure2:Experimental Analysis

In figure2 represent final hierarchy construction of Concepts where each link with arrows prerequisite relationship of concept and the value present on arrow define dependency of concept one concept to other concept.

concept	A-B	A-C	B-F	C-D	C-E	D-E	E-G	D-G	D-F
relationship	1	1	1	1	1	1	1	1	1

Build the final learning hierarchy

In concept learning hierarchy concept relationship is assign value 1 between concept and average correct prerequisite relationship is given above after complete experimental analysis. After complete computation the relationships between concepts is define in form of numeric value.

Conclusion

In this work our aim is to find the important content and their relationship for reading, using fuzzy logic . The representation of concept in form of hierarchy makes learning easy. We proposed work to extract important concept with very less time. The matrix representation is used for defining relationship between concepts. Experimental outcome is hierarchical representation of concept which show the dependency and the prerequisite of the concept.

REFERENCES

[1] Data Anonymization Approaches for Data Sets Using Map Reduce on Cloud: A Survey Veena1 , Devidas2 1 NMAM Institute of Technology, Nitte, Karnataka, India.
[2] A Solution For Privacy Protection In MapReduce Quang Tran,Hiroyuki Sato Graduate School of Engineering, The University of Tokyo.
[3] Travis Mayberry, Erik-Oliver Blass, Agnes Hui Chan, "PIRMAP: Efficient Private Information Retrieval for MapReduce", Proceedings of Financial Cryptography and Data Security (FC'13), pp. 371—385, Okinawa, Japan.
[4] M 2R: Enabling Stronger Privacy in MapReduce Computation Tien Tuan Anh Dinh, Prateek Saxena, Ee-Chien Chang, Beng Chin Ooi, Chunwang Zhang School of Computing, National University of Singapore.
[5] Airavat: Security and Privacy for MapReduce Indrajit Roy Srinath T.V. Setty Ann Kilzer Vitaly Shmatikov Emmett Witchel ,The University of Texas at Austin.
[6] https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html#Overview.