The Effect of Using the Semantic Mapping Strategy on the Reading Comprehension for the Eighth Grade Students in Al-Ramleh Elementary School for Girls in Zarqa District

Lina Abed Al- Gader Salameh

Abstract

Reading skill has not been given due care by the teachers of English language in the classrooms so students are no more lifelong readers. As a result students should be aware of the strategies that they can follow when they start reading in order for the reading process to be fruitful. One of the reading strategies that can help students comprehend what they read is the semantic mapping strategies. The researcher conducted a study to investigate the effect of using the semantic mapping strategies on the students’ reading comprehension. Therefore, the researcher conducted her research to answer the following questions:

1-Is there a statically significant difference between the mean scores attained by the experimental group on the pre-test and the post- test that can be attributed to the effect of using the semantic mapping strategy.
2-Is there a statistically significant difference between the mean scores attained by the experimental group and those attained by the control group on reading comprehension on the post-test that can be attributed to the semantic mapping strategies.

To answer the questions of the research, the researcher selected two groups randomly: The experimental and the control groups with fifteen subjects each. The two groups set for pre-test to assure that the two groups are equivalent. The researcher transferred three reading comprehension lessons into three types of maps: Concept map, word map and a story map. The experimental group was taught by the researcher using the three semantic maps to teach them reading comprehension lessons. On the other hand, the control group was taught by the traditional way which is illustrated in the teacher’s book.

The findings revealed that there are significant differences between the mean scores attained by the experimental group and those attained by the control group in favor of the experimental group due to the semantic mapping strategies; therefore, the researcher rejected the null hypothesis and accepted the alternative one which indicates that there is a significant differences between the experimental and the control groups. The researcher recommends in light of the findings that the students should be encouraged to follow the semantic mapping strategy while reading and to engage students in such activities that allow them to transfer more and more lessons into different types of maps such as word maps, concept maps and story maps.

Chapter One

Background of the Study

The ultimate goal of teaching English in Jordan is to enable students to communicate. Communication does not only take place with two or more speakers, but also it takes place between a speaker and a text. There should be a kind of interaction between the learner (reader) and the text which is being read in order to be good readers.

Reading is a necessary skill that any learner needs. Unfortunately, how to teach reading has not been given due care in our schools. In the past, according to the traditional view, reading begins with the child’s mastering the names of the letters, then mastering the letter-sound relationships, then learning some easy words in isolation, and finally reading simple stories with highly controlled vocabularies (Harp, B and Brewer, J 1996). Researchers and teachers as well complain that most learners are not able to understand what they read.

The teacher’s notion of reading is very important to determine the most suitable strategies and methods for reading effectively; it can also help in the way teachers tend to teach reading texts. In the past, teachers used to present a subject in the textbook and ask students to read whether silently or loudly, and then
students had to answer the questions that follow. Students, naturally, had no choice but to read even if they had not technical ways of how to read. What reinforced this perception of having any interest were the teachers' traditional techniques for teaching reading comprehension.

The result is that students hate to read, they only read the required textbook in order to be able to set for the achievement routine exams. In such case, students lacked motivation to read, even if they read, they show negative attitudes. For most of the learners, reading is an extremely difficult task that requires integrated body of skills, which also does not get easier with the passage of time and the accumulation of experience.

It is widely believed, however, that teaching reading is more difficult than teaching other language skills. Hence, to read effectively, the learner should be able to use two competences simultaneously: the linguistic and rhetorical. The latter is concerned with the style-system in cultural patterns of language. It is the rhetorical competence that most second language learners lack; unfortunately, having controlled the linguistic competence is not expected to read effectively or meaningfully. Therefore, students should be taught how to read in order to create lifelong readers (Harp, B and Brewer, J 1996).

Learner should probably develop reading comprehension skill much like learning to drive. Reading comprehension becomes so automatic that most skilled readers forget that they had to develop their reading comprehension skill. Learning reading comprehension requires a strategy where lesson plans progressively develop and reinforce reading comprehension skill, but a student does not seem to really get it by reading; this means that the student is successfully decoding words, but decoding without reading comprehension will not get him far.

So, what is Reading Comprehension? Reading comprehension skills separates the "passive" unskilled reader from the "active" readers. Skilled readers do not just read, they interact with the text. Skilled readers, for instance: Predict what will happen next in a story using clues presented in text, create questions about the main idea, message, or plot of the text, monitor understanding of the sequence, context, or characters (Sanders, M. 2001).

All teachers want their students to be good readers, but not all Agree on the best ways to teach reading. Effective reading requires not only accurate reading skills, but also to be able to comprehend easily and automatically (Lyon, 2001).

Many students who struggle to learn how to read are able, with appropriate instruction, to compensate for initial reading problems by becoming accurate decoders, but fail to reach a level of sufficient fluency to become fast and efficient readers (Adams, 1990). Further systematic research is needed to give us more comprehensive answers to questions concerning the best teaching reading strategies.

Learners comprehend better when they see the text organized in such a way which can easily be understood, and which indicates the relationships between ideas. One of the ways that may have a significant impact on the teaching reading process is the semantic mapping strategy.

Semantic mapping strategy can be used for at least several different instructional purposes. They can assist teachers in planning for instruction by helping them identify the patterns of organization of ideas and the concepts. A semantic mapping strategy can be useful for introducing the important vocabulary in a selection to be read. It shows students how the terms are interrelated. Teachers can use a semantic mapping to activate and tap student's background knowledge. Also, it can be a helpful reference for students to use in clarifying confusing points as they are reading. Once students are familiar with the nature of the semantic mapping strategy, they can create their own as a during-reading or post-reading activity.

Steps in the creation of semantic mapping strategy particularly the concept and word mapping strategies are: Analyze the concepts and vocabulary in the text. Arrange the words in a map that depicts the interrelationships between the concepts. Add to the diagram the words or concepts that are already understood by the students in order to depict the relationships between what they know and the information in the text (Chall, J. 1996). The semantic mapping strategy or Structured Overview, as it is sometimes called is a schematic diagram of the major concepts in a portion of text. The researcher of this study uses the semantic mapping strategy because this strategy can easily be taught and implemented by the students in addition to its significant role in developing students' thinking skills and reading comprehension. The purpose of this study is to apply the semantic mapping strategy, particularly, the word mapping, the concept mapping and the story mapping in transferring three reading comprehension texts taken from the Action Pack of the Student's Book of the Eighth grade and teaching students the maps created by the researcher to measure the effect of using these semantic maps on the students' reading comprehension.

Statement of the Problem
Most of Jordanian students do not know how to read meaningfully. They only read the textbook required to be able to perform well in the achievement tests. Students are reluctant to read English texts other than their school textbooks. They do not learn the processes for reading.

In spite of this, most of them get low marks in reading comprehension exercises. The problem may due to the way they used to follow while reading. They read the text as if it is consisted of discrete elements. Students do not interact with the passage they read, nor they build relationships between the terms in the text to build up the meaning, and then to lead themselves toward reading comprehension. Students are not aware of the strategies that may help them in reading, because they are not taught to do so nor are they immersed in reading activities to follow such strategies.

The main concern of this study is to determine the effect of the semantic mapping reading strategy on the students' reading comprehension, and to teach students how to use this strategy in reading through teaching them how to build up the structure of the maps. It is widely believed that learners learn better if they are taught to build up relations between the terms in such a text. Semantic mapping strategy is based on building up new relationships between the components of the text. (Mayer, R.2003) discussed the use of semantic mapping strategy. He defined the semantic mapping technique as being used to motivate and involve students in the thinking, reading, and writing aspects. It enhances vocabulary development by helping student link new information with previous experience.

Accordingly, the researcher believes that it is necessary to use semantic mapping strategy for teaching reading comprehension. Students need to read about what they are interested in and to interact and communicate effectively.

The Significance of the Study
The study aims at measuring the effect of using the semantic mapping strategies on the students' reading comprehension. The researcher uses three types of semantic mapping strategies; they are: Word mapping, concept mapping and story mapping.

The techniques of teaching language skills developed from using the teacher-centered technique into the learner-centered approach. Students need their teachers as facilitators and coordinators to improve their reading in an interactive and enjoyable manner (Clark-Edmands, S. 1998).

The purposes of the study are:
1- Using the semantic mapping technique to change three lessons taken from the Action Pack of the Eighth grade into three semantic maps: word map, concept map and story map.
2- Teaching students the semantic maps in an effective way.
3- Testing the effect of using the semantic mapping technique on students' reading comprehension.

The researcher uses three types of maps as a kind of variation not as a comparative study. She uses them to let students be aware of more than one type of semantic maps and for the students to be able to use the semantic map which suits for the type of reading comprehension text that they read.

The Questions of the Study
The study aims at answering the following questions:
1- Is there a statistically significant difference between the mean scores attained by the experimental group on the pre-test and the post-test that can be attributed to the effect of using the semantic mapping strategy.
2- Is there a statistically significant difference between the mean scores attained by the experimental group and those attained by the control group on reading comprehension on the post-test that can be attributed to the semantic mapping strategy.

The Hypotheses of the Study
1- There is no significant difference at ($\alpha=.05$) between the mean scores attained by the experimental group on the pre-test and the post test that can be attributed to the semantic mapping strategy.
2- There is no significant difference at ($\alpha=.05$) between the mean scores attained by the experimental group and those attained by the control group on reading comprehension that can be attributed to the semantic mapping strategies.

Limitations of the Study
1- As the study will be conducted on two classes of the Eighth grade students in a public school called Al-Ramleh Elementary School for Girls in Zarqa District in the second semester of the academic year 2006/2007, the results can not generalized beyond its population.
2. There are several strategies of the semantic mapping. In this study, only three strategies will be implemented: Word mapping, concept mapping and story mapping.

3. The scope of the study will also be restricted to females and to the procedures of carrying the study out.

4. The instrument of the study is the pre and post achievement tests for both: The experimental and the control group.

5. The implementation of the study will last only for two weeks because of the short specified time.

6. The study will be limited to certain types of reading comprehension; that is to say, three lessons of reading comprehension of the Action Pack of the Eighth grade will be transferred into maps and taught to the eighth grade female students.

Definitions of Terms

Semantic Mapping: It is a visual strategy which shows the major ideas of a certain topic and how they are related (Raymond C. Jones, 2006). In this study, word mapping, concept mapping and story mapping are used in teaching reading to display the interrelationships among ideas, words and the components of the story.

Word Mapping: An effective method, by which students enhance their understanding of key words by graphically mapping them. (see Appendix one)

Concept Mapping: It is a way to organize information about a problem or subject. It is consisted of nodes and labeled lines. Nodes are usually depicted with circles drawn around the term or concept. And the lines between nodes show which concepts are related (see Appendix Two).

Story Mapping: is a visual representation of the logical sequence of events in a narrative text. The elements of characters, setting, major events, problem, theme...etc. (see Appendix six)

Reading Strategy: The processes by which individuals analyze given information in a contextually specific situation and create new ideas and concepts based on their reading (Marier, R.2000). In this study, the strategy which has been used is the semantic mapping, more specifically, word mapping, concept mapping and story mapping

Traditional Strategy: the strategy in which the teacher presents the new vocabulary items first, and then he asks them to read the reading passage silently. After that, they answer the questions that follow.

Reading Comprehension: As defined by Partnership for Reading (2005), reading comprehension is understanding a text that is being read, or the process of “constructing meaning” from a text. Comprehension is a “construction process” because it involves all of the elements of the reading process working together as a text is read to create a representation of the text in the reader’s mind. In this study, it means that to what extent students comprehend the interrelationships between the ideas presented in the reading comprehension texts by the semantic mapping strategy and to what extent students can create similar maps as done by the researcher.

Chapter Two
Review of Related Literature

This chapter is divided into two main sections: theoretical and practical. Theoretically, the researcher will attempt to shed light on the meaning of semantic mapping in addition to the meaning of the reading comprehension, some examples of semantic mapping strategies, how can students reading comprehension be increased by using this strategy and the shared role between teachers and students in using this strategy.

On the practical level, the study summaries the results of relevant research studies. So, the study helps to direct the attention of English language teachers in general and the English language teachers of the elementary stage, in particular, to the significant role of the semantic mapping strategies in teaching reading in the English textbooks, particularly, the Action Pack textbooks, and perceive how students interact effectively with this strategy.
**Theoretical Section**

This review of literature covers studies related to strategies of semantic mapping. (Harvey, et al. 2000) mentioned that semantic mapping strategies are valuable instructional tools. Unlike many tools that just have one purpose, semantic mapping is flexible and endless in application. One common trait found among semantic mapping strategy is that they show the order and completeness of a student’s thought process - strengths and weaknesses of understanding become clearly evident. Many semantic maps show different aspects of an issue in close and also the big picture, since many semantic maps use short words or phrases, they are ideal for many types of learners, including English Language readers with intermediate proficiency. Tree maps can be used to show classifications, analysis, structures, attributes, examples, and brainstorming.

Raymond C. Jones, (2006) added that semantic mapping can be a helpful reference for students to use in clarifying confusing points as they are reading. Once students are familiar with the nature of the semantic maps, they can create their own as a during-reading or post-reading activity.

Williams, C. R. (1994) once said that semantic mapping enables students not only to visualize relationships, but to categorize them as well. As a direct teaching strategy that includes brainstorming and teacher-led discussions, it provides opportunities for schema development and enhancement, as well as prediction, hypothesizing and verification of content when used as a pre-reading activity. It is also referred to as a web or concept map. The teacher can introduce semantic maps to the class in different appearances. They can be shown as circles, squares, or ovals with connecting lines. The students read an assigned text. Through class discussion, the teacher writes the main idea of the text in the middle of the top circle. The students share the supporting details of the main idea and place them in circles that are connected to the main idea by lines. This activity can also be used by students in cooperative groups or individually.

Thomas, H. Estes (1999) explained that semantic mapping is a strategy for graphically representing concepts. Semantic maps portray the schematic relations that compose a concept. It assumes that there are multiple relations between a concept and the knowledge that is associated with the concept. Thus, for any concept there are at least three types of associations:

1. **Associations of class;** the order of things the concept falls into.
2. **Associations of property;** the attributes that define the concept.

He continued that the major purpose of the semantic map is to allow students to organize their prior knowledge into these formal relations, and thus to provide themselves a basis for understanding what they are about to read and study. Comprehension can be thought of as the elaboration and refinement of prior knowledge. What the semantic map provides is a graphic structure of that knowledge to be used as the basis for organizing new ideas as they are understood.

Heimlich, J. E., & Pittelman, S. V. (1986) explained that a semantic Map is one type of graphic organizer. It helps students visually organize and graphically show the relationship between one piece of information and another. This strategy has been identified by researchers as an excellent technique for increasing vocabulary and improving reading comprehension. As a pre reading activity, semantic mapping can be used to activate prior knowledge and to introduce key vocabulary words. As a post reading activity, words, categories, and new concepts can be added to the original maps to enhance understanding.

**Practical Section**

Zaid, M. (1995) applied the semantic mapping technique in teaching reading to his students at Abha college of Education. He explained that semantic mapping has been shown to be a beneficial learning/teaching technique for native speakers of English at all grade levels in regular and remedial classrooms as well as for those who are learning-disabled. He added that students who use semantic mapping manifest considerable improvement in reading comprehension, written expression, and vocabulary development. He suggested some areas of correlation between what a semantic mapping activity does and the principles and objectives of communicative language teaching (CLT). For the students, the map was providing a graphic conceptualization of their randomly given ideas. There are three places in a lesson where semantic mapping may be used as he clarified:

1. As a pre-assignment strategy to activate students’ prior knowledge or to help the teacher in assessing the students’ readiness to do the assignment.
2. As a strategy to allow students to record what they are learning during the assignment.
3. As a post-assignment strategy to allow them to integrate or synthesize what they have studied. He concluded that semantic mapping is interactive, it allows for sequential negotiation. It is an information-gap activity since students must fill in gaps in the map and in their personal schemata of the topic as the map takes shape. Moreover, it is a predictive activity. It is student centered because the semantic map makes use of the students’ prior knowledge and because students control the input at each stage of the map’s building. It is teacher-friendly because it allows the EFL teacher unobtrusively to pre-assess the students’ readiness to do an assignment, take immediate steps to enhance their preparation and to post-evaluate how well the students integrated or
graphically displayed their ideas in a semantic map. The authors described the multi-purpose usefulness of semantic mapping in the classroom. They evaluated the advantageous learning experiences for teachers and students asking students to brainstorm the ideas, images, or descriptions they associate with a particular concept. Semantic mapping and brainstorming programs were conducted to improve the students' reading and writing abilities.

Saqqa (2005) investigated the effect of computer assisted semantic mapping and brainstorming on Jordanian upper basic stage students' reading comprehension and writing in English. The findings revealed that students were very active, they read the texts from their textbooks, and then suggested some changes like deletion and additions on the first semantic map they drew. The researcher recommended that more computer assisted semantic mapping and brainstorming programs to be conducted to improve the students' reading and writing abilities.

Darayseh (2003) explored the effect of a proposed program based on semantic mapping and brainstorming strategies on developing the first scientific secondary students' English writing and reading ability. The findings of the study indicated that there were significant differences between the mean scores of the students in the experimental groups which can be attributed to the use of the semantic mapping teaching strategy. The researcher recommended that teachers should activate the prewriting phase and reading by using appropriate teaching strategies such as brainstorming and semantic mapping in particular.

Kuo and et al (2002) investigated the effect of concept mapping to enhance reading comprehension and summarization. They designed three concept mapping approaches: Map correction, Scaffold fading and Map generalization to determine their effects on the readers' comprehension and summarization ability. The experimental results of 126 fifth grader showed that the map correction method enhances reading comprehension and summarization abilities and that the scaffold fading method facilitates summarization ability.

Grigaite (2005) conducted a study to investigate the effect of using semantic mapping strategies on developing child's thinking skills. She defined semantic mapping as a strategy in which information is categorically structured in a graphic/visual representation. She examined the cognitive outcomes stimulated by the teachers' use of semantic mapping as a strategy for accelerating two cognitive operations, classification and seriation in a child's seventh year. Fifty-seven children at the age of six took part in the research. The findings revealed that students in the experimental group who participated in the training were creative. They revealed high degrees of cognitivism.

Schlesinger, C and et al (2000) elaborated on their beneficial experiences from incorporating semantic maps into class lessons, and the students recognizable academic improvement that resulted from utilizing this new teaching strategy. The authors quoted an eleventh grade student who reflects on her growth of knowledge while comparing her pre- and post-semantic maps she created on a six-week unit on Africa. The authors depicted semantic mapping as a graphic representation or picture of one's thoughts, ideas, and attitudes toward a key concept. Semantic mapping focuses on categorizing and connecting these thoughts, ideas, and attitudes in relation to the key concept. The authors detailed the process of semantic mapping as starting with teachers asking the students to brainstorm the ideas, images, or descriptions they associate with a particular concept, then students group related terms into categories, providing a label for each category. Then students graphically displayed their ideas in a semantic map. The authors described the multi-purpose usefulness of semantic mapping in the classroom. They evaluated the advantageous learning experiences for teachers and students through developing maps, either by individual students, or small groups or by the class. The authors then discussed the purposes semantic maps serve at different times during a unit. The authors concluded that the process of mapping techniques may improve reading comprehension, increase content-area achievement, enhance recall of material, and reduce student anxiety.

Kuo and et al (2002) investigated the effect of concept mapping to enhance reading comprehension and summarization. They designed three concept mapping approaches: Map correction, Scaffold fading and Map generalization to determine their effects on the readers' comprehension and summarization ability. The experimental results of 126 fifth grader showed that the map correction method enhances reading comprehension and summarization abilities and that the scaffold fading method facilitates summarization ability.

Canas, J.D et al (2004) conducted a study about text concept mapping, the contribution of mapping characteristics to learning from texts. The effects of text concept mapping were tested during one school year (4 classes, 112 eighth graders: two classes were taught using concept mapping with practicing. The other two classes were taught through regular learning skills). The classes were tested on language mapping comprehension after the teaching process. The findings indicated an advantage of using text concept mapping on reading comprehension. The researchers recommended that text concept mapping is a potent mediator for learning with texts and for conducting complex learning tasks, compared with concept mapping only.

Al Koumy (1999) compared the effectiveness of three classroom methods for teaching semantic mapping to college-level learners of English as a foreign language (EFL). Subjects were 187 freshmen at an Egyptian university; they were randomly assigned to three treatment groups: teacher-initiated semantic mapping; student-mediated semantic mapping; and teacher-student interactive semantic mapping. Treatment was administered over 5 months in one session per week. Subjects were pre- and post tested in reading comprehension. While the pretest indicated no significant differences in the groups, post test results revealed students in the teacher-student interactive semantic mapping group scored significantly higher than the other two groups, which had similar results.
Roy, B et al (2006) conducted a study to confirm a computer-based approach that can be used to score concept maps and then describe the concurrent criterion-related validity of these scores. The results indicated that automatically derived concept map scores can provide a relatively low-cost, easy to use, and easy to interpret measures of students' science content knowledge.

In conclusion, many studies revealed almost the same findings; they showed that the effects of using the semantic mapping do not only improve the learner's reading comprehension, but also their thinking, brainstorming and writing abilities.

Chapter Three
Methodology and Procedures

Chapter three deals with:
- Data collection procedures
- The population and subjects, design, instruments, validity and reliability of the instruments.

The population of the Study

The sample of the study consists of two classes of Eighth grade female students with thirty five students each in Al-Ramleh elementary school for girls in Zarqa district. The two classes were selected randomly. The researcher selected randomly fifteen female students from each class; fifteen female students to represent the experimental group and fifteen female students to represent the control group.

The Design of the Study

The design of the study is a complete experimental design. It consisted of two groups: The experimental and the control with pre-test and post-test for each group.

It promotes both teacher and learner in taking part in the teaching learning process. The teacher will give the two groups the pre-test. The control group will be taught the three reading comprehension lessons taken from the Action Pack of the Eighth grade in a traditional way while the experimental group will be taught the same reading comprehension lessons by using the semantic mapping strategies by teaching the three maps (see Appendix one, two and six). Then both groups will be given the post-test to investigate the effects of both strategies. Students are expected to be able to create new maps following the teacher's steps.

The Instrument of the Study

1-An achievement reading comprehension test which is to be used as a pre-test to both: the experimental and the control group.
2- An achievement reading comprehension test which is to be used as a post-test to both: the experimental and the control group with some modification related to the ways each group is taught accordingly.
3- A proposed three semantic maps (word map, concept map and story map) and activities for learning reading comprehension done by the researcher.

Procedures of the Study

The study will go to preparing the reading pre-test through the following procedures:
1- Preparing the reading pre-test to test the two groups.
2- Testing the two groups, correcting their papers, and calculating the mean scores to the experimental as well as the control groups.
3- Preparing the sample lesson plans, semantic maps and activities.
4- Establishing the required validity and reliability of the instrument.
5- Dividing the classes to be given for two weeks, six classes per week.
6- Planning the traditional way for teaching the control group. In the traditional way, the researcher teaches the control group the reading comprehension lessons in the traditional way which is illustrated in the teacher's book. The other three classes are used to teach the experimental group the semantic maps of the same comprehension lessons. In the second week, three classes are given to immerse the experimental as well as the control group in activities suit for the way each group was taught accordingly. The teacher then asks the students who belong to the experimental group to create similar maps to the other lessons in the same way. Then the two groups are given the post-test to measure the effects of using the two teaching reading strategies to calculate the two groups achievement scores and mean scores.
7- The traditional way is presenting the topic of the questions, asking students who belong to the control group to read silently and sometimes loudly, then asking students to answer the questions that follow each reading comprehension lesson. Next, students are asked to do the exercises in the work book.

8- The semantic mapping strategy is changing the three reading comprehension lessons into concept map, word map and story map by relating each idea to the other next one (see appendix one two and six...), then teaching students who belong to the experimental group the maps created by the researcher. Students then asked to create similar maps to the next reading comprehension lessons.

The procedures that the researcher followed when teaching each map:

Procedures of the Story Map
1- The teacher selects the story from the Action Pack of Eighth grade which is entitled "The Boy from the Past".
2- The researcher prepares questions to lead students through the story map, for example 'where did the story take place?' 'When did the story take place?'...etc.
3- The teacher discusses the organization of a story by explaining that any story has a beginning, middle and an end.
4- The teacher explains the visual story map (see Appendix six) and relates to story organization.
5- The researcher asks students to read the story and to fill out the map but without looking on the ready-made story map done by the researcher.
6- The researcher asks students about other stories they learned and asks them to compare them to the story they have been taught by the story map.

Procedures of the Concept Map
1- The researcher selects "Be Creative" reading comprehension lesson from the Action Pack of the Eighth grade.
2- The researcher constructs a map through nodes and lines bond through establishing relationships between the concepts.
3- The researcher clarifies the way the concepts are constructed in a map. And teaches them how to read the concept map done by the researcher (see Appendix two).
4- The researcher prepares questions to ask students about the concepts illustrated on the map.
5- The researcher asks students their attitudes toward this reading strategy. And ask them to create similar to the map they were taught by the researcher.

Procedures of the word map
1- The researcher selects "The Famous Scientist or Inventor" reading comprehension lesson.
2- The researcher built a word map to teach students the word 'scientist' through teaching them about one of the greatest Arab Muslim scientist called Alhazen. The researcher follows the following steps when constructing the word map:
   1- The researcher selects a word or concept central to the topic, which is 'Scientist'.
   2- The researcher displays the target word or concept such as 'scientist'.
   3- The researcher generates as many words as possible that relate to the target word such as investigator, discoverer... etc (see Appendix one).
6- The researcher leads the class in a discussion about the word map
7- The researcher then asks students to create similar word maps on the other next lessons.

Validity and the Reliability of the Instrument

Validity of the Pre-and Post-test
To determine the validity of this research instrument, both of the pre and post tests are given to a group of judges. The juries are two teachers with ten years experience each in teaching the basic and the secondary stages with B.A and higher Diploma in English teaching methods. In addition, the test is given to a supervisor with six years experience with M. A. degree. Also it was given to a teacher with M.A. degree
in English teaching methods. They all suggest some changes in the post-test to be able to investigate the
differences. The researcher modifies the post-test accordingly.

**The Validity of the Semantic Maps**

To investigate the validity of the semantic maps created by the researcher, the researcher follows some
models of concept maps which were done by a well-known researchers whose studies are trusted and
published in a well-known data base, such as EBSCO and ERIC data bases. Moreover, the semantic maps
are given to two supervisors with six years experience. They enrich the maps with some related ideas which
in turn will have significant role on the students' reading comprehension.

**The Reliability of the test**

In order to ensure the reliability of the pre- and post-test of the study, the researcher administers a test-
retest on an experimental sample of fifteen subjects who are taken from the population. Their tests are
corrected. A statistical formula of Cronbach Alpha is calculated. The result is (.909) which is a very high
internal contingency coefficient. This is a very appropriate for the reliability of the test and for the purpose of
the study as shown in the following table.

**Correlation is significant at the 0.01 level (2-tailed).**

<table>
<thead>
<tr>
<th></th>
<th>PRE</th>
<th>POST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>-</td>
<td>.909**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

**Chapter Four**

**Findings of the study**

The main aim of the study is to investigate the effect of using the semantic mapping strategies on the
reading comprehension for the Eighth grade students. To achieve this purpose, the researcher selects randomly two classes from (Al-Ramleh elementary school for girls) with thirty five students each. The researcher selected randomly fifteen students from each class. The first fifteen students are considered as an experimental group which is taught by the semantic mapping strategies and the other fifteen students are treated as a control group which is taught by the traditional way that is clarified in the teacher's book. Each group consists of 15 students of the Eighth grade.

**Findings related to the First Questions**

The first question of this study is: Is there a statically significant difference between the mean scores attained by the experimental group on the pre-test and the post-test that can be attributed to the effect of using the semantic mapping strategies.

To obtain data related to this question, the researcher administrated the reading achievement pre-test in the second semester of the academic year 2006/2007 on Wednesday, 15 of April at (Al-Ramleh elementary school for girls), just before the beginning of the experiment. Both the experimental and control group sat for the same format of the pre-test.

(Table one).

**Means and standard deviations obtained by the experimental group and the control group on the pre-test**

<table>
<thead>
<tr>
<th>GROUP</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE</td>
<td>15</td>
<td>3.53</td>
<td>2.356</td>
<td>.824</td>
<td>28</td>
<td>.417</td>
</tr>
<tr>
<td>experimental group</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In this table, the data collected on the pre-test are represented in the reading achievement—mean scores attained by the control group which is 2.87 while the experimental group is 3.53 on the pre-test. Using the t-test for dependent samples, the calculated t-value (.824) shows insignificance at (\(\alpha=.05\)) which indicates that the two groups are equivalent in their reading achievement; therefore, the researcher premeditated to use the t-test to work out all the comparisons regarding the two groups' total reading achievement scores on the post-test.

**The Findings Attained on the Reading Achievement Post-Test**

The second question is: is there a statistically significant difference between the mean scores attained by the experimental group and the those attained by the control group on reading comprehension on the post-test that can be attributed to the semantic mapping strategy.

### Table (Two)

Means and standard deviations obtained by the control and the experimental groups on post-test.

<table>
<thead>
<tr>
<th>GROUP</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Semantic mapping</td>
<td>15</td>
<td>12.07</td>
<td>2.052</td>
<td>11.353</td>
<td>28</td>
<td>.000</td>
</tr>
<tr>
<td>Control</td>
<td>15</td>
<td>3.33</td>
<td>2.160</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Teaching the semantic strategies to the experimental group lasted for two weeks, in addition to teaching the control group the traditional way.

Upon finishing the experiment, the experimental and the control groups sat again for the same test which is referred to as the post-test. The data collected from the post-test are represented in **(Table two)**. The table illustrates quite clearly the mean scores the students in the control and experimental groups get on the post-test. The total mean score achieved by the experimental group is 12.07 while the total mean score achieved by the control group is 3.33.

**(Table Three):**

Means scores obtained by the experimental group on the semantic strategies of reading on the post-test

<table>
<thead>
<tr>
<th>Experimental</th>
<th>Mean scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story map</td>
<td>15.2</td>
</tr>
<tr>
<td>Concept map</td>
<td>13.5</td>
</tr>
<tr>
<td>Word map</td>
<td>14.3</td>
</tr>
</tbody>
</table>

In this table, the researcher calculated the mean scores of the subjects of the experimental group on the semantic mapping different strategies to investigate if there are any significant differences between them. The researcher found out that the mean scores of three types of semantic mapping used by the researcher to teach students are almost the same.
Chapter Five
Discussion, Conclusion and Recommendations

The Final chapter aims at presenting summary of the results and discussions of the findings attained by implementing the semantic maps designed by the researcher who tried to transfer three lessons taken from the Action Pack of Eighth grade into three different maps: Story map to "The Boy from the past" , word map to "The Famous scientist or Inventor" and concept map to "Be Creative" . The second section includes the recommendations based on the findings.

The researcher uses three types of maps as a kind of variation not as comparative study. She uses them to let students be aware of more than one type of semantic maps and for the students to be able to use the semantic map which suits for the type of reading comprehension that they are reading.

Discussion of Findings

The findings of the study indicate that there are statistically significant differences between the experimental group and the control group. The mean scores of the students in the experimental group on the reading comprehension achievement post-test is 12.07 . It is considered higher than the mean scores achieved by the students in the control group which is only 3.33. The t-value calculated on the t-test is 11.353 showing a significant difference at (α=.05) as illustrated in Table two . Consequently, the null hypothesis "That there is no statistically significant differences at (α=.05) between the experimental and the control group which can be attributed to the semantic mapping strategies" is rejected while the alternative one " that there is statistically significant differences at (α=.05) between the experimental and the control group which can be attributed to the semantic mapping strategies" is failed to reject . Similarly, Al koumy (1999) investigated that the relatively high mean scores of the experimental group is evidently due to the effect of the implementation of the semantic mapping strategies.

One major purpose for story mapping is to assist teachers in planning and conducting reading instruction. Therefore, in preparing to have students read narrative selection, it is recommended that teachers analyze the structure of the story and create a map. The process of creating such a map helps teachers determine what is important enough about a story to be emphasized in class . For example, the theme often indicates background knowledge that students will need to use to comprehend, and this can become the focus of a pre-reading discussion. (See Appendix three).

The researcher observes how enthusiastic students were with this way of teaching reading. They were very active to the extent that they started giving their suggestions to develop the story map. One of their suggestions is to include the good things and the bad things of the characters and how they should behave in the story (see Appendix five). Their suggestions were unexpected and great. This proves what Raymond (2006) once said that once students become familiar with the nature of the semantic maps, they can create their own during-reading or post-reading activity. The mean scores of the students' achievements on the story map is 15.21 which reflects the improvement of the student's reading comprehension.

Moreover, using story maps to create questions will guide the discussion of the story; this discussion in turn will enhance students' understanding if the order of the questions posed by the teacher follows the organization of the story map. Also, consistently discussing stories in their logical sequence will strengthen students' sense of the important story grammar elements, and thus increase their ability to comprehend stories they will read in the future. The following are the questions suggested by the researcher that can be asked about each of the story elements:

1- Where does the story take place?
2- When does the story take place?
3- Could the setting have been different?
4- Why do you think the author chose this setting?

The researcher used to design the pre- and the post-test accordingly. Student's achievement scores are very high on the post-test questions related to the story map.

Utilizing the word mapping strategy is useful for helping students develop a general concept of definition. It makes them aware of the types of information which make up a definition and how the information is organized. When the researcher taught the lesson which is entitled " The Famous Scientist or Inventor") , most of the students become aware of the real definition of the word 'scientist' through teaching them one of the greatest figures of scientists in the Islamic Arab world and through teaching them his real achievements (see Appendix one). One of the students raised her hand and volunteered to give a definition of the word "scientist". She said that "scientist is a well-known and famous one who creates, investigates, and discovers great and new things". It was something wonderful from a student of Eighth grade to give such definition; this
indicates that students interact with the way they were taught by; which is the semantic mapping, particularly, word mapping as suggested by Heimlich (1986) that this strategy has been identified by researchers as an excellent technique for increasing vocabulary and improving reading comprehension. And as Zaid (1995) explored that students who use semantic mapping manifest considerable improvement in reading comprehension, written expression, and vocabulary development. The mean score of the students' achievements on the word mapping question is 14.3 which also indicates how word mapping is fruitful.

The third strategy of semantic mapping used by the researcher is the concept mapping. It is a good way to organize information about a problem or subject. Construction of concept maps helps us pull together information we already know about a subject and understand new information as clarified by Schlesinger et al. (2000) who depicted semantic mapping as a “graphic representation or picture of one’s thoughts, ideas, and attitudes toward a key concept.” They also added that semantic mapping focuses on categorizing and connecting these thoughts, ideas, and attitudes in relation to the key concept. The concept maps consist of nodes and labeled lines. Node is the name for important terms or concepts. Nodes are usually depicted with circles drawn around the term or concept, such as the node for "Be creative" "ability" (see appendix Two). The lines between nodes show which concepts are related. The label on the line "is" "to" as in the word map created by the researcher tells how or in what way the concepts are related. For example, "Creativity is the ability to invent, to imagine" The students at the beginning showed little interest when they were taught by the concept mapping strategy. Later, when the students were informed how to read the concept maps, they showed great enthusiasm toward it because they feel that they can build similar maps as built by the researcher depending on the steps they were taught and their background knowledge and their schemata as indicated by Thomas (1999) in reference to how semantic maps portray the schematic relations of the readers that compose a concept and how it assumes that there are multiple relations between a concept and the knowledge that is associated with the concept. Moreover, William (1994) explained that direct teaching strategy that includes brainstorming and teacher-led discussions provides opportunities for schema development and enhancement, as well as prediction, hypothesizing and verification of content as a pre-reading activity. The mean score of this strategy on the students' achievement post-test is 13.5 which is significant.

In conclusion, all the researchers who have conducted their researches on semantic mapping strategy agreed upon the significant role of applying the semantic mapping strategy in teaching reading comprehension for its great benefits that may serve a variety of learning purposes. They may serve as a student’s journal or record or instruction, providing students with a systematic means to integrate their new knowledge with their prior understanding, activating students prior knowledge and stimulating them to use that knowledge to interact with the text and promoting semantic mapping as a pre-reading activity that encourages students to map out their ideas.

Recommendations of the Study

In light of the findings of this study, the researcher puts forward the following recommendations; they are expected to be taken into consideration by the English teachers of Action Pack, the material producers and the curriculum designers:
1-The researcher recommends that the textbooks which are going to be adopted at Jordanian schools should include some guidelines that help students in the reading process.
2-Students should be fully engaged in the reading process: Pre-reading activities, drawing semantic maps...etc.
3-Encouraging students to gather ideas and develop them in a well organized concept maps.
4-Students should be encouraged to read for authentic purpose even if they are asked to read about certain topic chosen by the teacher.
5-Reading needs time and this time is necessary for students to grow, develop and improve their reading comprehension. Improvement entails practice. Time allocated for any reading task should be limited and agreed on for students to know how to read in an effective way using the semantic mapping strategy to plan to their reading.
6-Students need an environment that promotes reading and provides opportunities for individuals to work together in creating such maps. Teachers can to some degrees provide students with some technical ways of how to organize the reading process.
7-Training students to use the innovative reading strategies and techniques in their classroom such as the semantic strategy.
8-Students need modeling for teaching reading by using the semantic strategy ; therefore, teachers should familiarize students with the semantic mapping strategy by providing them with some models of maps.
9-Teaching vocabulary using the word mapping strategy.
10- Encouraging students to read stories and make maps following the general story map as illustrated in figure one and two.
11- Inclusion of semantic mapping activities in the technical repertoire of CLT as Zaid (1995) once recommended.

References
- El-Koumy, Abdel Salam(1999) The Effects of Three Semantic Mapping Strategies on EFL Students' Reading Comprehension, College Freshmen; College Instruction.
Appendix one
Word map

Inventor
Physicist

contributor

Scientist

creator
Pioneer
light
vision

Discoverer
Light goes into eye

Author
Kitab Al-Manazir

Study of eye

light

Investigator

The world of light

mirror

tenctions

Rainbow

Shadows

Reflections

Light travels in straight lines

Light goes into eye

Drawing accurate diagram of the human eye

Appendix two (Concept map)
Appendix Six (Story Map)
Done by the Researcher

- Be creative
  - ability
  - Ability
  - Ways of being creative
    - Finding a new use for things
      - In the past
      - Nowadays
      - Plastic
    - Changing how we do things
    - Evolution of ideas
      - Gradual
      - Dramatic
      - Send small tube into person's body

- Changing approach
  - Used for
    - Photographic films
    - Putting ideas together
      - Computer
      - Network
      - Internet

- Invent
- Imagine
- Solve problem
- Work on ideas

- Be creative
- Ways of being creative
- Evolution of ideas
- Changing how we do things
- Finding a new use for things
- Photographic films
- Putting ideas together
- Computer
- Network
- Internet

- In the past
- Nowadays
- Plastic

- Appendix Six (Story Map)
  Done by the Researcher
Abu Ali Al-Hassan Ibn Al-Haytham, or Alhazen, was born in Basra, Iraq in 965 AD, and died in 1040. He was probably the greatest physicist of his time. He investigated the world of light, including mirrors, reflections, rainbows and shadows.

Among other things, he discovered that light goes into the eye and that light travels in straight lines. He drew the first accurate diagram of the human eye.

He wrote over 90 works, including a study of the eye and vision called Kitab A-Manazir. In 1270, it was translated into Latin. It was published as a book in the west in 1572. Many later scientist used Alhazen's work.

General Story Map
(The Boy from the Past)
Action Pack 8

Setting: When: In the middle of the school days

Where: Uncle Abbas' house, the well, restaurant, garden, the land of the treasure

Characters: Yousef and Laila (sister and a brother), Abbas (Uncle), Mariam (Aunt), Ibrahim (gardener), Harry Dark (the dangerous man), professor Najjar (old friend of the family), thin man, the bald man, the short man (thieves of land treasures), grandfather (of Yousef and Laila), detective (the one who arrest the thief), Amer (a boy from the past).

Initiating event: Discovering an old Phoenician stone and a map in the well of the garden in Abbas' house.

Problem: The thieves attempted to steal the old valuable stone and the map to get the land treasure. They kept watching the house in the evening to steal them both.

Major Events: Laila and her brother Yousef surfed the net asking for a help to know about the old stone. They found unexpected e-mail from Harry Dark. They met Harry Dark who tried to take the old stone claiming that he is interested in the old things, but Laila refused. In one day evening, Laila saw a man watching the house, she hide the stone in the well. Laila and Yousef asked their uncle's help. They noticed directions on the map, they follow the map instructions and discovered the land treasure. Harry was following them to steal the land treasure, but he was arrested by the detective.

Resolution: Laila and Yousef decided to give the land treasure as a gift to a museum. And to keep their land treasures.
Perhaps the most important contribution he made was the modern scientific method. He created theories which he tested with experiments. This is how scientists still work today.

What does the word “scientist” mean?
A- Who is Al-hazen?
B- What did he investigate?
C- Was he the first one to discover that light goes into the eyes?
D- Mention some of his writings.
E- Write down the sentence which means that Alhazen’s book was very important one.

Question Number Two
Creativity
Evolution of ideas: Car development is an example of this. If a car designer creates a more comfortable car, he builds on all changes made by designers in the past.
Putting ideas together can be combined to create a new idea. If you join the ideas of a computer and a network, you get the Internet.
Changing how we do things: Sometimes new ideas bring about a change to how we do things. In the past all surgical operations involved cutting the patient. But if you send a small tube with a camera in a person’s body, you can treat him from the inside.
Finding a new use for things: If you look at something that exists, you can think of another use for it. For example the first plastic was developed for making balls, but now for photographic film.
Changing approach helps us find other solutions to the problem.

A- What does ‘creativity’ mean?
B- Mention the ways that can help us be creative?
C- May the thing have more than one use? Give example and write it on the sun shape.
D- Compare between the past and nowadays in terms of surgical operations?
E- How does changing approach help us be creative?

Question Number Three
Laila and her brother Yousef surfed the net asking for help to know about the old stone. They found an unexpected email from Harry Dark. They met Harry Dark who tried to take the old stone claiming that he is interested in the old things, but Laila refused. In one day evening, Laila saw a man watching the house, she hide the stone in the well. Laila and Yousef asked their uncle’s help. They followed the map, they followed the directions on the map to the land treasure. Harry was following them to steal the land treasure, but he was arrested by the detective.

1- Who are the characters that have been mentioned in the text?
2- Who arrested Harry Dark?
3- What are the major event that has occurred to Laila and Yousef?
4- Describe Harry Dark?
5- Where did the story take place?

Appendix Eight (Post-test)

Read the following text carefully then answer the questions in the space provided.
Abu Ali Al- Hassan Ibn Al-Haytham, or Alhazen, was born in Basra, Iraq in 965 AD, and died in 1040. He was probably the greatest physicist of his time. He investigated the world of light, including mirrors, reflections, rainbows and shadows.
Among other things, he discovered that light enters our eyes and that light travels in straight lines. He drew the first accurate diagram of the human eye.
He wrote over 90 works, including a study of the eye and vision called Kitab A-Manazir. In 1270, it was translated into Latin. It was published as a book in the west in 1572. Many later scientists used Alhazen’s work.
Perhaps the most important contribution he made was the modern scientific method. He created theories which he tested with experiments. This is how scientists still work today.

What does the word "scientist" mean?
- F- Who is Al-hazen?
- G- What did he investigate?
- H- Was he the first one to discover that light goes into the eyes?
- I- Mention some of his writings.
- J- Write down the sentence which means that Alhazen’s book was very important one.

Question Number Two

Creativity is the ability to invent or imagine something new. There are many ways of being creative and creativity can help us solve problems.

Evolution of ideas: Car development is an example of this. If a car designer creates a more comfortable car, he builds on all changes made by designers in the past.

Putting ideas together can be combined to create a new idea. If you join the ideas of a computer and a network, you get the Internet.

Changing how we do things: Sometimes new ideas bring about a change to how we do things. In the past all surgical operations involved cutting the patient. But if you send a small tube with a camera in a person’s body, you can treat him from the inside.

Finding a new use for things: If you look at something that exists, you can think of another use for it. For example the first plastic was developed for making balls, but now for photographic film.

Changing approach helps us find other solutions to the problem.

F- What does 'creativity' mean?
G- Mention the ways that can help us be creative?
H- May the thing have more than one use? Give example and write it on the sun shape.
I- Compare between the past and nowadays in terms of surgical operations?
J- How does changing approach help us be creative?

Question Number Three

Read the following story map then answer the questions that follow:

The Boy from the Past

Action Pack 8

Setting: When: In the middle of the school days

Where: Uncle Abbas' house, the well, restaurant, garden, the land of the treasure

Characters: Yousef and Laila (sister and a brother), Abbas (Uncle), Mariam (Aunt), Ibrahim (gardener), Harry Dark (the dangerous man), professor Najjar (old friend of the family), thin man, the bald man, the short man (thieves of land treasures), grandfather (of Yousef and Laila), detective (the one who arrest the thief), Amer (a boy from the past).

Initiating event: Discovering an old Phoenician stone and a map in the well of the garden in Abbas' house.
Indexed Semantic Mapping Rules

Haoning Zhang

(zhanghao@student.ethz.ch)

Master Thesis in Computer Science

March 2009

Supervised by

Prof. Donald Kossmann

Martin Hentschel

Tim Kraska

Contents

Abstract .................................................................................................................. 4

1

Introduction .......................................................................................................... 5

1.1

Motivation ........................................................................................................... 5

1.2

Project Goals ..................................................................................................... 5

1.3

Outline ............................................................................................................... 7

2

Fundamentals .................................................................................................. 8

2.1

XML & XQuery ............................................................................................... 8

2.1.1

XML Basics ................................................................................................... 8

2.1.2

Querying XML .............................................................................................. 8

2.2

Data Integration .............................................................................................. 9

2.2.1

Mapping Queries to Data ............................................................................... 9

2.2.2

Mapping Data to Queries ............................................................................... 10

2.3

Tree-based Index ........................................................................................... 10

3

B+ Tree based Rule-Aware Index ..................................................................... 12

3.1

Index Structure .............................................................................................. 12

3.1.1

Dependency Graph ........................................................................................ 13

3.1.2

Cycle Detection ............................................................................................... 14

3.1.3

B+ Tree with Insert Reference Counting ..................................................... 17

3.1.4

Rule Path Table ............................................................................................. 18

3.2

Semantics ........................................................................................................ 19

3.2.1

Add “is-a” Rules ........................................................................................... 19
Abstract

The Extensible Markup Language (XML), typically represented as text with explicit structure, is widely used as the standard in Internet, Data Exchange, and other technology domains. To manage different sources of information, several XML data integration techniques have emerged in last several years. The two most accepted solutions are information hub, which maps the data to queries, and query rewrite technique. One recent research project at ETH Zurich about Mapping Data to Queries (MDQ) provides a different approach, which introduces mapping rules to reflect data relationships and transform the data on-the-fly.
This master thesis presents an extended index structure that enables highly efficient and scalable XML query in existing MDQ engine. The key contributions of the work include: (1) Using index techniques from relational database to apply semantic mapping rules to large amounts of XML data; (2) Algorithms based on Dependency Graph to detect index impact when data or mapping rules are changed; (3) An implementation of that index as prototype that cooperates tightly with the existing MDQ engine.

In the experiments, the target solution shows significant query performance to process large amounts of data, when the mapping rules are fairly stable. The insertion or removing XML data or mapping rules apply the synchronization to the index transparently and with affordable cost.

1 Introduction

1.1 Motivation

XML as semi-structured data dominates data management and data exchange in these years. When the XML based data dramatically increases, the data management and integration problems are becoming more critical. XML data are fundamentally different from those in the traditional database management system in terms of their structure. In most institutions, enterprises, and within the whole Internet, XML data are kept in separated systems, where the data relationships are distributed maintained. As a consequence, data integration is widely accepted and used technology to manage those data. To describe the relationship between data, the Mapping Data to Queries (MDQ) [4] solution provides a schema mapping technique to describe subset relationships.

By defining the mapping rules as the input of the MDQ engine, the solution evaluates related results by its data relationships in a on-the-fly model, so theoretically it solved the data integration problem. However, in the practical situations, it will lead to performance penalties when the data volume reaches specific levels. The original query operation needs a full scan of all documents from the data source. When the size of XML data grows larger or data are stored in the distributed environment, the query performance are hardly bearable.

The root cause is the original MDQ engine is mainly designed for streaming processing, not for the database, where index structure are usually implemented. Without index support for a non-join query has to completely go through the data sources with n entities and costs $\Theta(n)$. In the current master thesis, the index mechanism is leveraged to support mapping rules in XQuery model.

1.2 Project Goals

How can we improve the performance when querying the XML data store with mapping rules?

In the traditional database, index techniques are used to increase the performance. Ideally, we would like to enjoy the benefits of the index mechanism to fasten the data mapping related query.

Existing index structures and algorithms obviously are not aware of mapping rules. Tree-based indexes and its derivatives like the B+ tree [2] are widely employed in most database systems, the keys indexed in the tree nodes come from the same table or entity, which does not fit crossing different data schemas because of data mapping relationships. The example in Figure 1-1 shows one common business case in international companies. Assuming the company has two branches, one in United States with employee information stored (e.g., monthly salary) in US Dollar. Similarly, the other branch in Switzerland maintains the personal records with salary stored in Swiss Francs. Moreover, both branches designed their own data formats. Now the request comes. The group CEO needs to find out the people in global scope, whose salary is between 5000 and 6000 Swiss Francs. Considering the exchange rate is 1.2 for CHF/USD, to fulfill the request, the request needs to be split for both indexes. In particular, for the first index, the query range is adjusted to (5000/1.2, 6000/1.2). Finally, both result sets need to be merged as the final result. Although this on-the-fly mapping algorithm works, the maintenance cost is not bearable when the number of entity classes and entity instances becomes larger. Moreover, the ever-increasing complicated mapping logic will kill the performance because the query
engine has to re-execute the mapping logics in each request. In Section 5.2 the experiment results show that the cost of naïve queries is linear to the size of XML data and number of schema mapping.

Figure 1-1 Data Integration

So the 1st goal of the master thesis is to propose one extended index structure to keep the intermediate mapping results. This idea is already implemented in traditional database as Index on

Rule: salary as $s is-a <gehalt>{$s * 1.2}</gehalt>

7 Indexed Semantic Mapping Rules

Materialized Views, to avoid computation effort when specific queries are frequently issued. When the mapping source entities are stored in target index with transformation, querying performance will be significantly increased because no mapping logic needs to be executed. However, as Materialized View in rational databases, storing pre-computed result leads to data consistency problems. In the former example, the index system needs to pay additional attention when (1) New employee entity comes from American branch, (2) exchange rate changed and it is not 1.2 anymore. In other words, the mapping rule is changed. So, the 2nd goal of the research is that index structure needs to be aware of rule mapping relationship and its consequence. Moreover, the index needs to be transparent to users when the data entity or mapping rules are updated.

In addition to the theoretical part, one prototype system is implemented to integrate with the current MDQ engine. With the system users can build the mapping-aware index on XML data nodes. In query processing sessions the corresponding index provides the MDQ engine tailored candidates to delimit the query range.

1.3 Outline

This thesis is organized as follows. Chapter 2 summarizes the background about XML, Mapping Data to Query (MDQ), and Tree-based indexes, which are the fundamentals of this work. Then the proposed index structure and its algorithms are illustrated in Chapter 3. Chapter 4 describes the implementation details of the prototype and how the new index integrates with existing MDQ query engine. My experimental results will be presented in Chapter 5. Chapter 6 concludes the work of the thesis with a brief discussion of ongoing work aimed at developing a more optimized index structure.

2 Fundamentals

2.1 XML & XQuery

2.1.1 XML Basics

The Extensible Markup Language (XML) is a general-purpose specification for creating custom markup languages [14]. It is widely used to describe data and its structure in many application domains, where users define their own mark-up elements. The self-describing property aids information systems to understand each other and share their data. The text-based format makes the XML easily pass through layers of the communication model or even cross different platforms.

Owning those advantages, XML becomes one standard of data storage format. Several XML database systems [10] [11] have emerged in recent years, some of them are dedicatedly using the XML data model, and others are extending relational database such as Oracle and DB2. In terms of the physical storage, one naive approach is to store the XML text in one table in the relational database, where the traditional advantages as security, transaction control, and index mechanism could be re-used. Another approach is to map the hierarchical XML data into multiple tables [12].

2.1.2 Querying XML

When the amount of XML documents dramatically increases, several problems emerge. The first problem is the query technique, in other words: how to find the right information
efficiently from huge amounts of data. In the world of structured data, information is stored as records in relational databases, which also provides a standard query language and interface (SQL) to users or its applications. The similar query mechanism must be provided before XML information could be used in large-scale information systems. The philosophy of XML is that the correlation of data and schema is not necessary, the data and schema relationship could be M:N.

XQuery [1] is a query language that is designed to solve the above problems; it provides the definition to access information in XML data. In the view of XQuery, the collection of XML documents will be accessed as tables in relational database. The standard is already implemented in several open source projects like Saxon [8] and Zobra [9]. In addition to retrieve the information, XQuery also supports to construct new XML elements or documents in its query result. This property make XQuery is very successful in content management, media, and web application domains.

2.2 Data Integration
When the size of unstructured data is greatly increased, the data integration problem becomes one of the most significant issues to solve. Data integration is the process of combining data residing at different sources and providing the user with a unified view of these data. For example, one typical automobile manufactory may require querying its thousands of vendors' inventory system to collect supply chain information. Each vendor uses its own data schema to describe inventory details. To provider a unified interface for users or upper level applications, scalable schema mapping technique is widely used to solve this problem. Because the query and data are not consistent with each other, two approaches could be involved in this integration process: either (1) Mapping queries to data, or (2) Mapping data to queries.

2.2.1 Mapping Queries to Data
The first approach is to query those information sources respectively, and then merge the result for host application. This mechanism provides an uniform query interface over a mediated schema. This query is then transformed into specialized queries to the original databases. Mapping queries to data is often used due to the simplicity. The central application aligns the queries against data sources, and those data sources are not necessary to change, which is critical in lots of business scenarios. This implementation is also called as Local As View (LAV) [13], while the local data sources are considered as database views. To map queries to data, several query rewrite techniques are proposed. Query rewrite algorithm captures the original query statement and splits it into a set of source queries to data sources. After successfully receiving all the result sets from data sources, the algorithm merges (union operation) the result set and sends it back to user. The idea of rewriting the query comes from the relational database domain (i.e. views). The drawback is that users have to rewrite the query (views in SQL) whenever a new data source is to be integrated or an existing source changes its schema. Moreover, the performance for querying against large number of diverse schemas is becoming more inefficient. The rewritten queries can become very large. In addition, the effort to manage the data dictionary is not bearable. So, this approach is not scalable, it is only suitable for small number of schemas.

2.2.2 Mapping Data to Queries
Instead of rewriting the query to meet different schema, the Mapping Data to Queries (MDQ) [3] [4] technique introduces an alternative approach through rule-driven mapping mechanism. The MDQ models the XML data as a Directed Acyclic Graph (DAG) to represent the data
mapping relationship. Each edge in the DAG represents a parent-child relationship. The “is-a” rule adds more edges to the graph, and those rules support transitive property. By this approach, several achievements are valuable in real scenarios: (1) the local data schemas are unchanged; (2) the effort to maintain the rules is obviously smaller than managing queries in query rewriting model; (3) the user applications are not aware of the schema diversity; and (4) the transformation is fired on-the-fly, which reduces the impact when updating the rules.

2.3 Tree-based Index
One critical feature of relational databases is to handle large amounts of data. In order to store and retrieve those data efficiently, index techniques are widely used. In relational databases the B+ tree [2] is widely used to represent sorted data and offer an efficient way to query, insert, or remove records by keys. The B+ tree extends the B tree to store all the data at the leaf level, and the interior nodes only store keys. Moreover, the B+ tree builds the bi-directional link chain at the leaf level, which makes the range query more efficient. The structure of B+ Tree is shown in Figure 2-1.

For a b-order B+ tree with h levels of index, performing a range query with k elements occurring within the range requires $\Theta(\log_b n)$ operations in the worst case, and the space required to store the tree is $\Theta(n)$. In this processing model, the query engine calculates the result on-the-fly. The advantage is that the original B+ Tree index can be reused with zero maintenance cost in the index part, but the drawback is also obvious: In the experiments introduced in Chapter 5, the system in this naive query model has a huge burden when the rule related queries are frequently requested. The alternative approach, which is introduced in this chapter, is to provide the index with a stable query performance of $\Theta(\log n)$, where $n$ represents the size of XML data. Instead of applying the mapping rules after receiving the query request, we pre-compute the rule mapping

Indexed Semantic Mapping Rules

For a b-order B+ tree with h levels of index, performing a range query with k elements occurring within the range requires $\Theta(\log_b n)$ operations in the worst case, and the space required to store the tree is $\Theta(n)$. In this processing model, the query engine calculates the result on-the-fly. The advantage is that the original B+ Tree index can be reused with zero maintenance cost in the index part, but the drawback is also obvious: In the experiments introduced in Chapter 5, the system in this naive query model has a huge burden when the rule related queries are frequently requested. The alternative approach, which is introduced in this chapter, is to provide the index with a stable query performance of $\Theta(\log n)$, where $n$ represents the size of XML data. Instead of applying the mapping rules after receiving the query request, we pre-compute the rule mapping
and transformation in advance and resemble the B+ Tree index. In this circumstance, when users request the XQuery like “/b”, the query engine directly access the index on b, without being aware of the rule logics. To achieve this goal, the index system must be aware of the mapping rules, and is designed to maintain the relationship between data and rules.

The proposed index structure, shown in Figure 3-1, contains three parts: (1) Dependency Graph, (2) B+ Tree with Reference Counting, and (3) Rule Path Table. They will be introduced in the next sections in this chapter.

### 3.1 Dependency Graph

The transitive properties of mapping rules makes the rules not standalone or independent. They may influence each other to represent more implicit rules. For example, the rule a is-a b and b is-a c create the implicit rule a is-a c. The implicit rule makes the index hard to maintain its data. As its mechanism, the indexes always get one “copy” from the main data and make it be synchronized when main data is updated. In the previous example, when “c” is indexed, this index need to be informed when “a” is updated. In the target solution, the dependency graph is used to dynamically probe the transitive paths and their implicit rules.

A dependency graph is a directed graph representing dependencies of several objects towards each other [15]. It is possible to derive an evaluation order or the absence of an evaluation order that respects the given dependencies from the dependency graph.

When we leverage the dependency graph to analyze transitive rules, the graph nodes can be seen as XML classes and edges as mapping rules. Figure 3-2 shows how the dependency graph works in mapping rules analysis. In the sample, the edge R1 represents a is-a b, and R2 represents b is-a c. Because of the transitive nature of the “Is-A” Rule, we can induce that a is-a c. The mechanism also produces more implicit rules like b is-a x and a is-a x.

### Rule Path Table

From the previous sample we notice the transform path for the second “emp” is <R1, R2, R1, R3>. The rule R1 shows two times because of the cycle. The mechanism of Automata provides us the traditional way to solve the problems like regular expression. However, in this MDQ case the automata solution is not so qualified. The first reason is, in the MDQ engine each mapping rule only fires once. In other words, the transform paths have limited steps (length). The second reason is that, it is useless to input duplicate rules into MDQ as parameter. The MDQ engine is able to detect cycles on-the-fly. The third reason is the irrelevance of the rule order. When MDQ reads the rule set, the order of the input rules are not critical.

Table. In the B+ tree, each leaf node exists with a positive reference count, which means at least one Leaf Reference points to it. So the result is, the total size of Leaf References is greater or equal than the size of its B+ tree.

When the mapping rules are fairly stable and not asked to be removed, the Rule Path Table is not necessary when user build the index. Even the remove is occasionally required, the Rule Path Table can be generated dynamically.

### 3.2 Semantics

The semantics of Rule-Aware Index are defined in the following sections. The target index
supports not only traditional index semantics like insert and remove, but also reflects changes when mapping rules are added or removed. In Chapter 3.2.1, the semantics of adding a “Is-A” rule is defined. Chapter 3.2.2 - 3.2.4 introduces the semantics of data insert, remove, and update. Chapter 3.2.5 defines the semantic of rule removal.

3.2.1 Add “is-a” Rules

Assuming node “a” and “b” in one Dependency Graph, when one new “a is-a b” Rule is introduced, only “a” and its predecessors are impacted. The impacts include: (1) some nodes could reach “b” and its successors because of the new rule; (2) some nodes are able to reach “b” and its successors by alternative paths.

We explain this theory with an example in Figure 3-6, which shows the Dependency Graph with a Terminating Node “z”. Assuming the R5 is the last rule added into the graph, then the impacted nodes are {m, x, s}. Note that although data from “s” and “x” can already reach “z”, R5 provides them an alternative path to the Terminating Node.

Figure 3-6 Mapping Rules in the Dependency Graph

4 Integration into MDQ

In this chapter the prototype system which implements the theory part of the previous chapter will be introduced. The prototype is implemented as an Eclipse RCP program in Java.

4.1 Architecture Overview

As shown in Figure 4-1, the whole system was designed in four parts: (1) Index Management, (2) Storage Management, (3) Interface to MDQ Engine, and (4) Query Interface.

Figure 4-1 Architecture

When user queries or manipulates the from the Query Interface, the request is first passed to the Index Management module, which analyzes the possibility to leverage indexes. In particular, when mapping rules are introduced or removed, the Index Management module checks the dependency graph to analyze the impact of existing indexes.

If a specific index is picked up for the user request, the related range query is executed in its B+ Tree structure and then one list of related XML file IDs are returned. Those file IDs are candidates for the query. Through the Storage Management part the system gets the file handles to the actual file system.

Once the XML file handles are ready, the MDQ Engine Interface part packs it to the MDQ Engine, combined with rules exported from the dependency graph. The execution of mapping query generates the output as XML results, and they are shown to users in the Query Interface.

4.2 Index Management

The Index Management module has two major parts. The first part is the pool of Rule-Aware Indexes, which includes one traditional B+ tree with reference counting, and one attached Rule Path Table per index. The second part is the Dependency Graph, which only exists once and is shared among all indexes. In the prototype system, the Catalog window, shown in Figure 4-2, manages XML data, index instances, and elements (nodes and rules) in the Dependency Graph.

Figure 4-2 Catalog Window

The structure and algorithm of Rule-Aware Index are fully described in Chapter 3. In addition, two design issues need to be solved. (1) Which indexes should be used during the query, and (2) Which indexes should be updated when mapping rules are changed (insert or remove).

The prototype solves the first problem in the manual approach. To pick up the right index automatically, the XQuery statement need to be analyzed, which is out of scope of the thesis. The manual way to assign the index is introduced in Chapter 4.5.

For the second problem, the following algorithm finds out all the impacted indexes when the mapping rule “a is-a b” is either inserted or removed:

Step 1 Calculate the successor node set s of in the Dependency Graph;
Step 2 Calculate the node set t, whose elements bind at least one index;
Step 3 Update the index:

Indexed Semantic Mapping Rules

```plaintext
39
forecah n in (s intersect t) {
forecah i in indexSet(n) {
i.update();
}```
4.3 Storage Management

In the tree-based index structure, the database system extracts the key range from the target query and then tries to find the tree leaves which hold the keys. In B+ tree leaves, the location of target information piece should be found.

In this solution for XML documents, instead of storing the location of XML nodes, the tree leaves store the location of XML file (absolute file path). The advantage of this approach is that we bypass the complexity of unstructured text context when building the index, and make more flexible to ad-hoc queries.

However, the absolute file path, which is stored as String type is still too large for B+ Tree leaves. So in the implementation all files are indexed in one array of strings, and the index of the array member is the File ID, which stores at the leaves of the B+ Tree. Figure 4-3 illustrates the storage structure.

4.4 MDQ Engine Interface

The MDQ Interface is taken as the single connect point to the MDQ Engine. The main tasks of the interface consist of (1) Issue query request against XML documents, and (2) Scan XML documents to extract index information for indexing.

4.4.1 Query XML Documents

As shown in Figure 4-4, the index management model communicate with MDQ engine through queryMDQ() function, which includes two variants. The former queries against the whole XML storage, and the latter only targets one specific XML data file.

Refer to Section 3.1.1, the index management model maintains a dependency graph, which is shared to all indexes to analyze the dependency relationship. In the graph each edge represents one mapping rule, so one specific path in this directional graph equals to one vector of mapping rules.

In the example in Figure 4-4, the node “a” can reach the terminating node “x” through maximally four paths \(<R1, R2, R5>, <R1, R4, R5>, <R3, R2, R5>, <R3, R4, R5>\). Particularly, to find out all the instance of “a” that could be mapped to “x” through the specific path \(<R3, R2, R5>\), the following function call retrieves the candidate set:

```
List candidates = queryMDQ(<R3, R2, R5>, "//x")
```

Notice that the candidates result set candidates includes the data instance not only from “a”, but also possibly from “b”, “c”, and “x”. So, to get all the results from “a”, one more filter step need to be fulfilled.

Figure 4-4 Interface to MDQ

provides a easy-to-use environment for querying with index support. As shown in Figure 4-6, the Eclipse form allows user to input the XQuery statement and specify maximal two indexes and their key range. When two indexes are applied, the user need to clarify the AND/OR predicates according to the XQuery’s WHERE clause.

Once the query is successfully executed, the query result is shown in the middle column, and the execution cost and its index selectivity are listed on the right “History” table.
5 Experimental Evaluation

5.1 TPoX Benchmark

Transaction Processing over XML (TPoX) [6] [7] is an application-level XML database benchmark based on a financial institution scenario. Three major entities, Customer, Security and Order describe the simple security trading system. It is used to evaluate the comprehensive performance of XML database systems. In the current experiment environment we test the correctness and performance of the new index structure by using the documents generated from the TPoX Benchmark as the original scenarios. To evaluate the mapping rule aware scenarios, some of original data instances are rewritten.

The prototype system was implemented in Java and all experiments are performed on a PC-compatible machine, with a single Intel 2.5GHz T9300 CPU and 2GB RAM, running Windows Vista Enterprise (SP1) and JRE 1.6.0. Each experiment is repeated five times and the average of the five results is used as the measurement.

5.2 Query without Mapping Rules

In the TPoX benchmark, seven core queries are defined to measure the query performance. The first six queries only query against one schema, the seventh query which needs to join two different schemas is not supported by current implementation of the MDQ index. So in the final experiments, the following six core query were used.

For each core query, two configurations need to be evaluated. We first execute the XQuery against specific number of XML files without any index support, then in the second configuration we build the indexes and re-evaluate the query performance. The former configuration is called “Naïve Mode”, and the latter is shown as “Indexed Mode”. Table 5-1 and Figure 5-1 summarizes the execution results in our environment, and the conclusions are:

- In Naïve Mode, the costs are shown as linear curves. The reason is that, all the files related to the same schema need to be scanned, no matter of the selectivity of the predicates in the XQuery;
- The Indexed Mode show significant improvements, and the cost are almost stable. The underlying B+ Tree provides the performance of $\Theta(\log n)$, where $n$ represents the size of XML data.

5.3 Experiments with Mapping Rules

To evaluate the performance when mapping rules involve in XML data, we derivate the TPoX data source to generate 3 different data set, which include 10, 100, and 1000 schemas respectively. Meanwhile the mapping rule set for those schemas are provided.

5.3.1 Building Index

The cost to build the MDQ index with Rule Path Table is expensive. In the algorithm introduced in Chapter 3.2, each XML document need to be scanned for each Rule Path in the Rule Path Table of the MDQ index. So theoretically the cost model could be described as:

$$\Theta(\ast)$$

: the size of XML data, $\ast$: the number of rule paths

In the index without Rule Path Table, only one rule path that includes all the mapping rules is built. So in this case $\ast=1$, so we can deduce that the building cost for indexes without Rule Path Table is $\Theta(\ast)$, where

is the size of XML data.

In the experiments, the result proved the theory in the thesis. Table 5-2 to Table 5-4 illustrate the cost to index 10000 files is linear to the number of schema and size of XML data, when the algorithm take extra computation to track the rule-data relationships in Rule Path Table. The experiments also show the efficiency in building the index without the table. The absence of the table does not impact the correctness of query. In addition, it saves the memory space as well. However, as the penalty, the whole index has to be rebuilt when any of rules is updated or removed.
Richard C. Sinatra
Josephine Stahl-Gemake
David N. Berg

Using semantic maps (webs or networks) is a popular technique among reading teachers. The maps improve children's reading comprehension by showing them not only how vocabulary words are related to each other in some conceptual hierarchy, but also how the ideas in texts are organized in associative ways. Mapping is a cognitive strategy; it necessitates that the teacher conceptually organize the text for student use. As this article will show, particular map organizations and the relationships suggested through connective links express nonverbally the relationships of parts to the whole and the whole to its parts. A study of children in our reading clinic, using several types of semantic maps, has shown that the technique is effective with reading disabled youngsters. This study will also be described here.

A semantic map
A map or web is a graphic arrangement showing the major ideas and relationships in text or among word meanings. The map consists of "nodes" (drawn in circles, rectangles, or squares) containing key words or phrases, with connecting links (lines or arrows drawn between the nodes).

Symbols common to flow charts may be used. For instance, a circle generally means program start or finish, a triangle means that a decision is to be made, while the rectangle holds the operations expressed in the program.

As an example, in a map for an episodic tale, the nodes are linked together by arrows showing the flow of events in the story. The title is