

# Finding Hidden Semantics of Text Tables

Saleh A. Alrashed

Royal Saudi Air Force, Riyadh, Saudi Arabia  
Dralrashed@gmail.com

**Abstract.** Combining data from different sources for further automatic processing is often hindered by differences in the underlying semantics and representation. Therefore when linking information presented in documents in tabular form with data held in databases, it is important to determine as much information about the table and its content. Important information about the table data is often given in the text surrounding the table in that document. The table's creators cannot clarify all the semantics in the table itself therefore they use the table context or the text around it to give further information. These semantics are very useful when integrating and using this data, but are often difficult to detect automatically. We propose a solution to part of this problem based on a domain ontology. The input to our system is a document that contains tabular data and the system aims to find semantics in the document that are related to the tabular data. The output of our system is a set of detected semantics linked to the corresponding table. The system uses elements of semantic detection, semantic representation, and data integration. In this paper, we discuss the experiment used to evaluate the prototype system. We also discuss the different types of test, the experiment will perform. After using the system with the test data and gathering the results of these tests, we show the significant results in our experiment.

## 1 Introduction

Documents are one of the most important ways of sharing knowledge between humans. They are constructed using some common assumptions about their structure. Authors intend to convey information in ways allowing readers accurate and effective interpretation of the contents. This is why understanding documents is a relatively easy task for an intelligent human reader. One of the ways that authors use to present information in documents is tables.

The number of tables used per page in scientific papers has increased quite steadily over time. It has grown from 9% of the pages in 1984 up to 32% in 1997 [5]. Document tables have been created by humans to aid understanding of the information, therefore any attempts to reuse their content automatically needs effort to recognise and determine the table's structure and semantics.

Because of the large number of documents that have been published electronically, there is a real need to reuse the contents of these types of documents in investigations. This implies a need for automatic analysis of documents to

aid human users. As an important part of a document, tables have received attention from researchers trying to locate, analyse, identify and transform them into reusable formats for further analysis by software systems [3, 6, 7]. Most researchers have concentrated on the table itself, identifying its physical and logical structure, without considering the relation between a table and the surrounding text in the document [4].

The problem with isolating tabular data that appears in textual documents from the rest of the text, is that although the table data can be extracted and reused, it is not possible to fully understand its contents and reuse it effectively in other integration processes unless this data has been combined with parts of the text in the same document that are related to that table. This text describes the semantics of the information in the table.

This problem arises for two main reasons. Firstly, the author of the document tends to explain parts of the table in the text around it, and the information in the table does not make sense if the table is completely isolated (semantically) from its document. Even if there are no explanations about the table in the text, the table can not be completely isolated from its domain, especially if it is going to be integrated with other tables without losing some of its usefulness. Secondly, locating all data with related description of the semantics in the table structure makes the table difficult to understand as it extends the size of the table and affects the clarity of presentation. Therefore authors tend to leave parts of the meaning of data to be explained in the surrounding text.

## 2 Semantics and Representational Conflicts

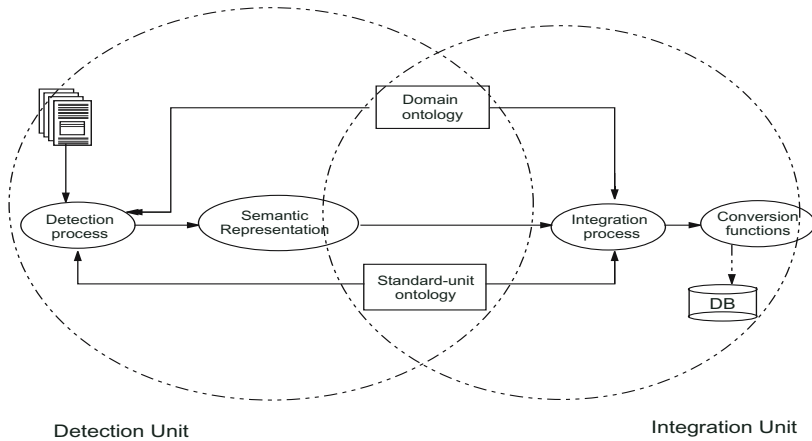
When trying to integrate data from different sources drawn from the same domain, semantic conflicts and representational conflicts can occur. Semantic conflicts are related to differences in the metadata of the table (attributes, names etc). Such a conflict occurs when different data names represent the same data (synonyms), the same name represents a different data domain (homonyms), or a hidden semantic relationship exists between two or more terminologies. For example the relationship between cost and price or between profit and net-profit, and the similarity between car, vehicle and truck cannot be understood unless we use a domain knowledge base to relate these terms. To overcome conflicts we use semantic metadata, which provides information about the meaning of the available data and any semantic relationships [1]. The sources of semantic information can include ontologies. Madnick [2] defines an ontology as an agreement about a shared conceptualization of a given application domain.

Representational conflicts occur due to the way data is represented and in particular the measurement unit being used. These conflicts are concerned with the values of an attribute. If we have two attributes, which are semantically the same, they are not in conflict when their values are represented in the same way, i.e. in the same units. This is important, if we are bringing attributes together. For example, the values of a price can be represented in pounds or in dollars. If the units are different, there is a representational conflict unless they can be

converted to the same representation before they are used together. Thus representational metadata provides information about the meaning of the values of an attribute, its representational relationships and its units of representation.

### 3 Semantic and Representation Detection Framework

SRD (Semantic and Representation Detection framework) is to be a prototype system for discovering and interpreting the context information about tables present in the text of a document containing tabular data, and prepare them for interoperation with other data. Figure. 1 shows the proposed system architecture of the SRD system, which will extract and structure the context data about a table held within a textual document.



**Fig. 1.** Semantic and representation detection framework architecture

It consists of two main units and each unit has a number of processes and sub-units. The first is the **Detection Unit**. The main purpose of this unit is identifying, extracting and representing the information about the context of a table's elements given within the text. This information will be used to enhance the table's metadata and thus lead to better use of its contents. It operates on the documents, which are represented in ASCII characters and have tabular data in them.

The second unit is the **Integration Unit**. This is concerned with integrating the detected semantics with other semantics to create a description of a table and its representation. Documents are created by different authors who represent the data using their knowledge of the domain. Therefore the detected semantics might need to be converted into a common representation and terminology before data from this table can be integrated with data from another. In this paper, we discuss the experiment used to evaluate the prototype system. We also discuss

the different types of test, the experiment will perform. After using the system with the test data and gathering the results of these tests, we show the significant results in our experiment.[9-11]

## 4 Experimental Design

In our experiment, we use documents from different sources available on the Internet. The documents were collected randomly, and had to contain tabular data to be valid for the experiment. We have used documents from four primary data sources, as this allows us to detect any differences due to a data source as well. We will investigate the differences between commercial data and non-commercial data sources, and between scientific domains and non-scientific domains with the four types of data source.

In the experiments we have a number of input variables which describe the experiment. These are:

1. The number of domains used. In our experiment we use two types of domain- a scientific (chemistry) and a non-scientific (cars). By using different domains, we will see if there are differences in the number of detected semantics in documents representing different domains.

2. The number of data sources for the domain. For each domain, we used two sources of data. In the cars domain, we used data from the Imotors and Which web sites.

- Which web site: This is a consumer magazine Web site based in the UK. It gives independent, unbiased advice and evaluations on different products and services. This Web site issues a monthly report on different types of car. It reviews many aspects about the cars (e.g. performance, security, and price). Most of the reports contain tables which present the results of an evaluation. With each table, there is text data around it, which explains the table data and important issues about the car. We have extracted 50 pages from this web site. These were selected at random and are different evaluations published on different dates.( See [www.Which.co.uk](http://www.Which.co.uk) ).
- Imotors web site: This is a commercial website. It is one of the fastest, easiest ways to purchase a car online. It has a USA national network of car dealers who provide competitive quotes, and an inventory of thousands of used vehicles. Imotors is now a car-buying service focused on effectively matching consumers with the vehicles they want. Imotors web pages normally consist of a table, which lists the cars that meet the user requirements and a paragraph or two which describe some of the values in the table. We have extracted 100 pages from this web site, which represent different search results.( See [www.Imotors.com](http://www.Imotors.com) ).

In the scientific domain (chemistry), we used two web sites - the Thermoset web site and the Eastman web site.

- Thermoset is part of The Lord Corporation and traces its roots to 1919. Lord's ideas produced inventions, and led to chemical formulations, bonding

processes, elastomers, adhesives, coatings, bonded elastomer assemblies and many more discoveries. They provide on their web site, descriptions and specifications of their products. We have extracted 100 pages from this web site. This website is a commercial website. ( See [www.thermoset.com](http://www.thermoset.com) ).

- Eastman Chemical Company (NYSE:EMN) is a global company which is one of the world's largest suppliers of polyester plastics for packaging. Headquartered in Kingsport, Tennessee, USA, Eastman manufactures and markets more than 1,200 chemicals, fibres and plastic products. On their web site, they evaluate a number of their chemical products so that customers can determine whether a product meets their requirements. We have extracted 50 pages from this web site( See [www.Eastman.com](http://www.Eastman.com) ).

It is probable desirably to use other data sources to determine whether they show the same behaviour. However there was insufficient project time to identify another source domain with enough documents meeting our needs for a specific domain. Thus, for the two domains chosen we have two different types of sources- a commercial site ( Imotors and Thermoset ) and an evaluation source ( Which and Eastman ). Further work is needed to confirm the results hold in other domains.

3. The number of documents from each data source for the commercial data sources is 100 documents and 50 for the evaluation data sources, which makes a total of 300 documents. We have tried to increase the number of documents but consistent sources of these types of documents ( documents containing tabular data) are hard to find. The number of documents within each source type is sufficient for the experiment, because as Admantios mentions in [8], 30 items can show the true behaviour of an item with respect to its characteristics being evaluated.

#### 4.1 Experiment Objectives

For our experiments, we have a number of objectives we wish to evaluate. These are:

1. To show that there are hidden semantics - in documents - that are related to a table in that document and they are significant to its interpretation.
2. To show the usefulness of a domain ontology in detecting and using these semantics.
3. To determine the difference between alternative approaches that can be used to detect such semantics and to identify the best approach.
4. To determine whether there is a relationship between the number of semantics in a paragraph and the distance between the paragraph which contains the semantics and the table itself.
5. To identify whether there is a significant difference in the number of detected semantics in different domains.
6. To evaluate the system.

To achieve these objectives, our experiments will analyse the documents and calculate a number of values. These analyses and calculations will be performed

twice. The first time, we will perform them manually ( fully human, no use of any other system ). The second time we will use our system SRD without any human interference. Doing the analysis twice will allow us to compare the two methods and discover if there is a significant difference between a manual approach and the SRD approach. We then do comparisons with the SRD system to determine the effect of using an ontology in the determination of semantics and to identify where the semantics occur in the text in relationship to the table.

The values that are going to be calculated in this comparison are:

1. The number of semantics detected using the table header augmented by the appropriate terms from the domain ontology as the search key (enhanced keywords).
2. The number of semantics detected using the table information ( data, meta-data, footer).
3. The number of semantics detected in the paragraphs before the table.
4. The number of semantics detected in the paragraphs after the table.

We use the experimental data to perform the the coming tests.

#### **For Each Data Source**

1. Compare the number of semantics detected using the domain ontology to enhance the search keys with the number detected using only keys from the table information.
2. Compare the number of semantics detected in the paragraphs before the table with the number detected in the paragraphs after the table.
3. Compare the total number of semantics detected using our system (SRD) with the total number of semantics detected manually.

**For Each Domain.** Compare the number of semantics found for the two types of data sources in the domain.

1. Number of semantics in Imotors web pages with number in Which web pages, detected using the table keys enhanced by terms from the domain ontology.
2. Number of semantics in Imotors web pages with number in Which web pages, detected using table information only.
3. Number of semantics in Imotors web pages with number in Which web pages, detected using the paragraphs before the table.
4. Number of semantics in Imotors web pages with number in Which web pages, detected using the paragraphs after the table.
5. Total number of semantics in Imotors web pages with total number in Which web pages.

These tests will also be performed using the chemistry domain web sites when Thermoset web pages will be compared with the Eastman web pages. These tests will also be performed to compare the two sets of commercial web pages, and the two sets of scientific web pages, i.e. a comparison of the Imotors web pages with Thermoset web pages and the Which website with the Eastman website.

## 5 Experimental Results

We have divided the significant results that we found into two parts, namely tests related to the logical content of the documents, and tests related to the physical structure of the documents.

### 5.1 Tests Related to Logical Content of Documents

We found that there is a significant number of useful semantics, hidden in a document, which are related to the tabular data in that document. After analysing 300 documents from different data sources, we found that there is hidden data related to the tabular data in the document which can be detected, extracted and represented as beneficial semantics related to that table. These semantics can be used for integrating these tables with data from other data sources.

Our experiment has shown that the number of detected semantics is normally distributed among all the 300 documents with a mean of 9.9 semantics per document. This means that the average number of semantics detected in a document is 9.9. The experiment sample documents can be categorised by the name of the domain that the sample is related to. Also, within each domain there are two types of data sources, commercial and evaluative.

We have used two domain types, scientific ( chemistry ) and non-scientific (cars). We found that the chemistry domain had fewer semantics than the cars domain, with the chemistry mean equal to 8.3 and the cars domain mean equal to 11.5. Thus, the cars domain has a higher number of semantics per document, and we think that the reasons for this are as follows:

1. The chemistry domain has well-defined concepts, therefore the users or the writers of the documents do not have many alternative terms to identify and describe the concepts. In other words the domain name concepts don't have many synonyms or relationships and are well known to the community.
2. As a result of the first reason, the writers of a document don't have many things to explain in the text.
3. The cars domain has different terminologies between different countries and even between groups of people; therefore the writers of a document in this domain need to explain most of the attributes in their tables, as they are targeting a set of readers with a less coherent background.

We have shown that detecting the hidden semantics will result in a better understanding of a table, and in its enrichment with extra semantics.

**Significance Between Different Domains.** 80.7 % of the detected semantics have been detected using a domain ontology approach. In our experiments, we used two approaches to detect the hidden semantics. One of these approaches used a domain ontology. This approach gives us most of the semantics that were detected manually in the text, as shown on Table 1. This table, shows the percentage of semantics detected by the domain ontology approach in each data source of the total number of semantics that were detected manually. Thus, most

**Table 1.** Percentage of semantics detected in documents using domain ontology

Data source name	Domain name	Percentage
Imotors	Cars-Commercial	74.8
Thermoset	Chemistry-Commercial	87.9
Which	Cars-Evaluation	71
Eastman	Chemistry-Evaluation	89.2

of the semantics are detected using a domain ontology approach. Therefore, any event that affects the domain ontology will affect the total number of semantics detected, and any changes in the domain ontology will also affect the detection process. For instance, if the domain ontology becomes richer in concepts, then the total number of semantics that are detected will increase, and vice versa. This led us to conclude that a domain ontology plays an important role in detecting hidden semantics in documents, and any limitations in the domain ontology will also limit the number of semantics detected.

We also found that the number of semantics detected using a domain ontology approach in a non-scientific domain (cars) is higher than in the scientific domain (see table 2). But with respect to the percentage of the total number of semantics detected manually, the scientific domain has a higher percentage, as shown in Table 1.

**Table 2.** Number of semantics detected in each data source using domain ontology

Data source name	Domain name	number semantics
Imotors	Cars-Commercial	7.18
Thermoset	Chemistry-Commercial	5.53
Which	Cars-Evaluation	4.10
Eastman	Chemistry-Evaluation	2.82

**Significance of the Differences Between Data Sources.** There seems to be a significant difference between the number of semantics in the commercial and evaluation data sources in both our domains. Comparing the number of semantics detected in commercial data sources in the cars and chemistry domains has shown that they are always higher in the commercial data sources than in the evaluation data sources, see Table 3. However, this needs further investigation to determine if it holds in other domains.

We believe that the large difference between these types of data sources is due to a number of reasons:

1. Commercial data sources try to give the reader as much information as they can to attract him or her to buy their merchandise, whereas in evaluation data source documents they evaluate a product without looking to the need to attract the reader to purchase. This is in part, because they are showing facts which the supplier and the reader might not like.



**Table 3.** Number semantics for each data source

Data source name	Domain name	Data source type	number semantics
Imotors	Cars	Commercial	9.59
Thermoset	Chemistry	Commercial	6.29
Which	Cars	Evaluation	5.76
Eastman	Chemistry	Evaluation	3.16

- Commercial data sources have to attract the consumers to buy their products using a small space. Therefore, they put all the information they have into a very short text which makes it a quicker to read and also easier to detect by other systems, whereas in evaluation data sources, the time constraint is not a crucial element as their readers want as full a comparison as possible.
- The length of a document has an effect on the number of semantics. The evaluation data sources are longer documents than commercial data sources, yet have less table semantics. This needs more investigation, but may indicate they are giving a fuller evaluation in the text of the points being made.
- Also, most of the documents in the evaluation data sources describe new products and technologies which have not yet been included in the domain ontology. This may be a cause of the smaller number of semantics detected in them, but this needs further investigation.

**Significance in the Prototype System.** Our system SRD has detected 70 % of the total number of semantics detected manually in the sample documents. By comparing the total number of semantics detected by SRD with the total number of semantics detected manually, we found that for all data sources SRD detected a reasonable percentage of the total semantics, (see Table 4). This Table shows a significant difference between the number of semantics detected by SRD and manually, and we believe that the reason for this is not the method we are using, but two other reasons:

- There is a weakness in the programming and thus the keyword searching in our system is not as sophisticated as it might be. If we had used a better searching mechanism, we might have achieved better results. For example, when searching for engine size the system is able to detect " engine size " but not able to detect "engines-size" in the text.
- The domain ontology used is limited. For example, in one of the documents there was a semantic value in the text mentioning that 'the car has 260 HPs' but 'HP' was not a term in the domain ontology, and therefore the SRD system did not detect this semantic.

The high percentage of undetected semantics by the SRD when compared with the manually detected semantics has occurred because of the weakness of the domain ontology used in our system. We found that most of the missing semantics could be related by an ontology if the ontology was expanded. However, without a well defined ontology the missing semantics will remain undetected.

**Table 4.** Number of semantics detected by SRD and Manually

Data source	Domain name	SRD	Manually	Percentage
Imotors	Cars	9.59	12.38	77.5
Thermoset	Chemistry	6.29	9.09	70
Which	Cars	5.76	9.76	59.1
Eastman	Chemistry	3.16	6.8	46.5

We believe that this point needs more investigation using a richer ontology. However, it shows that an automatic detection system using an ontology is unlikely to achieve 100% due to the difficulty of getting a comprehensive ontology.

## 5.2 Experimental Significance Related to the Physical Structure of Documents

**Adjacent Paragraphs.** We found that 75.05 % of the total detected semantics come from the adjacent paragraphs to the table. In Table 5, the commercial data sources from both domains have a very high percentage of their semantics coming from the adjacent paragraphs, while in the evaluation data sources the percentages are not so high.

**Table 5.** Percentage of semantics in the adjacent paragraphs

Data source name	Domain name	Percentage
Imotors	Cars-Commercial	88.3
Thermoset	Chemistry-Commercial	86.1
Which	Cars-Evaluation	64.8
Eastman	Chemistry-Evaluation	61

We believe there are three reasons for this:

1. The commercial data sources are reasonably short compared to the evaluation documents, and this will affect the spread of the semantics in the paragraphs. In commercial data sources, documents have 2 to 7 paragraphs, whereas in evaluation data sources they can go up to 13 paragraphs.
2. Evaluation data sources tend to talk about one concept or part of the table in each paragraph and start on a new topic in a new paragraph. This leads to the semantics being spread over the document.
3. Commercial data sources try to concentrate the information into one or two paragraphs to hold the reader's attention.

**Paragraphs Under the Table.** We found that 70 % of the detected semantics appear to be from the paragraphs after the table. It is common that the writer of the document will describe the table after showing it to the reader. In our sample, the table in commercial data sources usually comes in the middle of the

**Table 6.** The mean of the number of semantics before and after a table

Data source name	Number of Semantics before Table	Number of Semantics after Table
Imotors	3.09	6.50
Thermoset	1.68	4.61
Which	1.98	3.78
Eastman	.76	2.40

document, whereas in evaluation data sources it is normally near the beginning of the documents.

In Figure 7, we can see that the mean of the number of semantics after a table is always higher than the mean of the number of semantics before the table for all data sources. In Imotors, the paragraphs after the table have produced 65% of the total semantics detected in the documents, also in Which they have produced 66%, Thermoset 74%, and in Eastman they have produced 76%. In total the paragraphs after the table give us 68% of the total number of semantics. This tells us that detecting semantics in the paragraphs after the table is more productive than using the paragraphs before the table.

We believe that by concentrating on the adjacent paragraphs and paragraphs under the table, a system will get most of the semantics, if not all of them presented in a document. In some cases, the tables are put at the end of the document, or in a certain place in the text, or refer to it by its number. We treat this type of document as if the table is in the first indicator position in the document text.

**Semantics and Indicators.** By tracing the types of indicators used in the text, we find that there are a number of types that the writers tend to use to point to the table from the text. The first one, and the most commonly used is indicating a table by its number, for example 'Table 2.1', and this appears when there is more than one table in the document. It is also used when the table is far away from the indicator in the document. The other types of indicators are 'the above table', 'the next table', 'the last table', 'the previous table', 'in the table below' and the use of 'Figure' instead of 'Table' in these phrases. These types of indicator always need to be close to the table. In some cases there is no indicator in the documents, and this is because the document is short and there was only one table.

In Table 7 P is Paragraph, S is Semantics, and I is Indicators. The column heading P with I and S means paragraphs with an indicator and semantics. This analysis shows that paragraphs containing indicators nearly always have semantics in them (Column 3) and that the semantics are always next to the indicator. Comparing between paragraphs, the paragraphs that have indicators have most of the detected semantics in them (Column 4). Therefore, in large documents, it is useful to search for the indicators first and concentrate on the paragraphs that contain them when looking for hidden semantics.

**Paragraphs and Indicators.** Among those paragraphs that have indicators, the first paragraph has the highest number of semantics in it. We believe that this

**Table 7.** Percentage of semantics in paragraphs that contain indicators

Data source	P with I	P with I and S	S in P containing I	S in the 1st P
Imotors	80	100	89.5	51.7
Thermoset	83.3	100	92	38.2
Which	90.9	95	94.7	10
Eastman	88	90	98.2	13.6

result needs more investigation, because the type of documents we used might affect this result. The documents we used are slightly short and might not show the real situation. However, in the evaluation data sources, the writers spread the semantics throughout the documents and sometimes they use different indicators, for example 'in the first column' and 'in the last row' appear in these documents showing a fuller analysis is being undertaken.

## 6 Conclusion

As an experiment, we applied our system SRD to 300 documents related to two domains, cars and chemistry. For each domain, we used documents from two different data sources, a commercial and an evaluation. The experiment has shown that documents from commercial data sources in both domains have more semantics than those from evaluation sources. Also, documents from the scientific domain, chemistry, have fewer semantics than documents from the non-scientific domain, cars.

We found from the experiments that it is common that the writer of the document will describe the table after showing it to the reader. We found that 75.05 % of the total detected semantics comes from the paragraphs adjacent to the table. Also, we found that 71.8 % of the detected semantics appear to be from the paragraphs after the table. We believe that by concentrating on the adjacent paragraphs and paragraphs after the table it is possible to get most of the semantics, if not all of them, that are present in the text.

We also found that paragraphs that have indicators always have semantics in them, which are always next to the indicator. Thus, when comparing between paragraphs, the paragraphs with indicators have most of the detected semantics. Therefore, in large documents, it may be useful to search for the indicators first and concentrate on the paragraphs that contain them in the full detection process. By tracing the types of indicators, we found that there are a number of ways that the writers tend to use to indicate to a table in the text. The first one and the most commonly used is indicating to the table by its number for example 'Table 2.1'. The other types of indicators are 'the above table', 'the next table', 'the last table', 'the previous table', 'in the table below' and with 'figure' instead of table. This can be useful in directing the searching mechanism to concentrate on paragraphs that contain these types of indicators. Also, in large paragraphs, the hidden semantic is sometimes close to that indicator.

## References

1. Christof Bornhovd (1999): "Semantic Metadata for the integration of Web-based for Electronic Commerce" International Workshop on Advanced Issues of E-Commerce and Web-based Information Systems (WECWIS'99) Santa Clara, California, 1999
2. Madnick , S.E. (1995): From VLDB to VMLDB (Very Many Large Database: Dealing with Large-Scale Semantic Heterogeneity, Proc. 21st VLBD Conf. 1995.
3. O. Hori and D. S. Doermann, Robust table-form structure analysis based on box-driven reasoning, Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1), 1995,218,IEEE Computer Society
4. Debashish Niyogi, A Knowledge-Based Approach to Deriving Logical Structure from Document Images, Department of Computer Science, SUNY Buffalo, 94-35, 1994,
5. Matthew Hurst, The Interpretation of Tables in Text, University of Edinburgh, 2000
6. Pallavi Pyreddy and W. Bruce Croft, TINTIN: A System for Retrieval in Text Tables, 2nd ACM International Conference on Digital Libraries, 1997,
7. M. Yoshida and K. Torisawa and J. Tsujii, A method to integrate tables of the World Wide Web, Proceedings of the First International Workshop on Web Document Analysis ,2001, 31-34, Seattle, Washington, ICDAR'01
8. Adamantis Diamantopoulos and Bodo Schlegelmilch , Taking the Fear Out of Data Analysis, 1997, The Dryden Press, London
9. Saleh Alrashed, W. A. Gray, (2002) "Detection Approaches for Table Semantics in Text", Proceedings of the 5th International Workshop, DAS2002, Princeton.
10. Saleh Alrashed and W. A. Gray, Semantic Detection for Tabular Data in Text, In 7th World Multiconference on Systemics, Cybernetics and Informatics, 2003, Orlando, Florida , USA, IEEE, Computer Society ,
11. Saleh Alrashed and W. A. Gray, Utilising Semantic Conversion Functions to Link Tabular Data, In The 9th. International Conference on Information Systems Analysis and Synthesis: ISAS '03, 2003, IEEE, Computer Society