# Object-Net Approach for Data Extraction and Reporting

**M.Munusamy[1], G.Tholkappia Arasu[2]**

**[1] Anna University-Coimbatore,
Tamil Nadu, India.**

**[2] Jayam College of Engineering and Technology,
Anna University-Coimbatore, Tamil Nadu, India.**

**Abstract**

This paper attempts to present an Object-Net method for understanding the meaning of a natural language (i.e. English), and generate the SQL query based on underrating, and produce the data report as per the requirement understanding. It is proposed to model the elementary meanings which assist the machine to autonomously undertake the analysis and synthesis processes of meaning. In the proposed methodology disambiguation process is performed in context manner: starting from natural text, the context of the sentence is identified, then the actual meaning is identified using correlation of elementary object meanings exist in Object-Net database, It is because even ambiguous word will have only one meaning based on the context or object or domain on which the sentence is written. The data warehouse (DW) is a database. The data stored in the warehouse is uploaded from the operational systems. Data warehouse provides the information required by the decision makers. Business intelligence (BI) mainly refers to computer-based techniques used in identifying, extracting, and analyzing business data, such as sales revenue by products and/or departments, or by associated costs and incomes. BI technologies provide historical, current and predictive views of business operations using the data from data warehouse (DW). The cost of building a data warehouse (DW) & business intelligence (BI) is expensive for any organization as it requires data warehouse tools for building data warehouse and extracting data using data mining tools from data warehouse. The proposed method called Object-Net uses the English language for getting the requirement for business intelligence reporting and identifies meaning of the sentence, internally creates the interface layer, generates query, gets the data and reports for analyzing the data.

***Keywords:*** *Context, Database, Domain, NLP- Natural language processing, Object, Parse tree, WSD- Word sense disambiguation, DW- Data Warehouse, BI- Business Intelligence.*

## 1. Introduction

Word sense disambiguation (WSD) is the process of identifying which sense of a meaning is used in any given sentence, when the word has a number of distinct senses [4]. For a long time the WSD is an open problem in natural language processing (NLP). The solution of this problem impacts other tasks such as discourse, engines, anaphora resolution, coherence, inference, information retrieval, machine translation and others. This paper attempts to present an object net method for data extraction from database and reporting using natural English language. It is proposed to model to autonomously undertake the analysis and synthesis processes of meaning the elementary meanings of English sentence and mapping it as per the actual database structure from which user requires extracting the data, and writing the SQL query for the input English sentence or the requirement which is written in regular English. It is different from existing approach on database query language like SQL, PLSQL, Oracle, Sybase, because any of the database engine can able understand only its own query language but if we try to enter our requirement interims of English sentence the database engine will not process our sentence and it will give error message. The proposed here an algorithm understands the English language and maps the equivalent corresponding database query language and in such a way that it can be used to interact with the database engine with its language to retrieve the data and then report it to user.

## 2. Word sense disambiguation (WSD) approaches, Data Warehouse (DW) and Business Intelligence (BI)

There are two main types of approach for WSD in natural language processing called as deep approaches and shallow approaches.

**Deep approaches:** these approaches involve the intention to understand and create meaning from what is being learned, Interact vigorously with the content, Make use of evidence, inquiry and evaluation, Take a broad view and relate ideas to one another, and Relate concepts to every time experience [3], [6], [9]. These approaches are not very successful in practice, mainly because such a body of knowledge does not exist in a computer-readable format, outside of very limited domains. There is a long tradition in computational linguistics, of trying such approaches in terms of coded knowledge and in some cases; it is hard to say clearly whether the knowledge involved is linguistic or world knowledge. The first attempt was that by Margaret Masterman, at the Cambridge Language Research Unit in England, in the 1950s, and Yarowsky's machine learning optimization of a thesaurus method in the 1990s.

**Shallow approaches:** These approaches are not concerned of learning the text instead they deal with the surrounding words of the ambiguous word and try to identify only parts of interest for a particular application. They just consider the surrounding words, using a training corpus of words tagged with their word senses the rules can be automatically derived by the computer. This approach, while theoretically not as powerful as deep approaches, gives superior results in practice, due to the computer's limited word knowledge.

In addition to deep approaches and shallow approaches, there are four conventional approaches to WSD:

**Dictionary and knowledge-based methods:** These approaches make use of dictionaries, thesauri, and lexical knowledge bases, without using any corpus evidence.

**Supervised methods:** These approaches make use of sense-annotated corpora already been trained from semantically disambiguated corpus.

**Semi-supervised or minimally-supervised methods:** These approaches make use of both labelled and unlabeled data for training - typically a small amount of labelled data with a large amount of unlabeled data [10].

**Unsupervised methods:** These eschew (almost) completely external information and work directly from raw corpora (i.e. not annotated).

**Data Warehouse:** A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process. A data warehouse maintains its functions in three layers: staging, integration, and access. Staging is used to store raw data for use by developers. The integration layer is used to integrate data and to have a level of abstraction from users. The access layer is for getting data out for users. The data from data warehouse is used for many purposes like, business professionals for data mining, online analytical processing, market research and decision support. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system.

**Business Intelligence:** It is usually refers to make decisions based on the information that is available in data warehouse. A data warehousing (or data mart) system is the backend, or the infrastructural, component for achieving business intelligence. Business intelligence also includes the insight gained from doing data mining analysis, as well as unstructured data. Business intelligence aims to support better business decision-making. Thus a BI system can be called a decision support system (DSS). Though the term business intelligence is sometimes used as a synonym for competitive intelligence, because they both support decision making, BI uses technologies, processes, and applications to analyze mostly internal, structured data and business processes while competitive intelligence gathers, analyzes and disseminates information with a topical focus on company competitors. Business intelligence is a subset of competitive intelligence.

The method proposed here is a semi-supervised method; it is called as object - net approach which uses the information dynamically gathered from user that is while machine finds any of untrained corpora or unable to solve the disambiguation then those information are reported to user or master, after user understand the problem the related corpora are trained [7]. It differs from previous semi-supervised approaches: the algorithm has a set of disambiguated trained elementary objects, and incrementally builds and resolves the untrained elementary objects. This algorithm can be incorporate into lager applications like machine translation, code generation, search engine, IR, etc.

**Resources:** The algorithm does not dependant on any other existing WSD resources like WordNet, SemCor, and any BI tools like SAP Business Object, . Instead of that it uses separate database named as Object-Net Database which contains trained elementary objects. Initially the database is stored with limited data, this database updated when new untrained object found in the input text or when fine tuning is required on existing already trained element. The proposed algorithm finds all its required information to identify the meaning of the word on a particular context from this Object-Net database, so precision of word sense disambiguation of proposed algorithm mainly depends on data from this special Object-Net database.

## 3. Object - Net Approach for Data Extraction and Reporting

3.1 Object –Net Approach Procedures: The algorithm presented in this paper determines, in a given text, a set of nouns and verbs which can be disambiguated with high precision, the semantic tagging is performed using the sense defined in Object-Net Database, and actual meaning of the sentence is identified. But above mentioned task are completed in step by step using methods, so the various methods used to identify the correct sense of a word are presented first, Next presents Object-Net Database architecture, the main algorithm in which these procedures are invoked in an iterative manner, and the method of updating, fine tuning the Object-Net Database.

PROCEDURE 1: This procedure tokenizes the given sentence and creates a parse tree path for the given sentence. Parse tree paths were used for semantic role labelling. Predicates are typically assumed to be specific target words (verbs), and arguments are assumed to be spans of words in the sentence that are dominated by nodes in the parse tree. A parse tree path can be described as a sequence of transitions up from the target word then down to the node that dominates the argument span. The parse tree paths are particularly interesting for automated semantic role labelling because they generalize well across syntactically similar sentences. For example, the parse tree path in fig 1 would still correctly identify the "taker" argument in the given sentence if the personal pronoun "she" were swapped with a markedly different noun phrase.

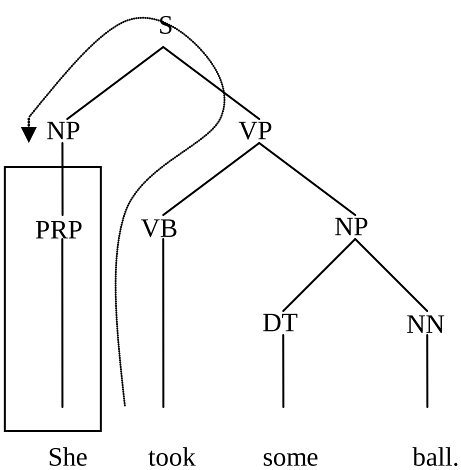


Fig 1: An example parse tree path from the predicate "took" to the argument "She", represented as ↑VB↑VP↑S↓NP

PROCEDURE 2: Identify the words having only one sense(monosemous words) in Object-Net database and make them as having number of sense as #1.
Example: the noun subcommittee has one sense defined Object-Net database. So this is a monosemous word and marked as having sense #1.
PROCEDURE 3: with this procedure, we are trying to get contextual clues regarding the usage of the sense of a word. For a given word Wi, at position i in the text, form two pairs, one with the word before Wi and the other one with the word after word Wi. Then we find out all the occurrences of these pairs found within the Object-Net database. If, in all the occurrences, the word Wi has only one sense as # Wis, then mark the word Wi as having sense #Wis.
PROCEDURE 4: Find the words which are semantically connected to the already disambiguated words for which the connection distance is 0. The semantic distance is computed based on the ObjectNet hierarchy. Two words semantically connected at a distance of zero if they belong to same path of subnet.
PROCEDURE 5: Find words which are semantically connected [2] in ObjectNet and for which the connection distance length is zero. In this procedure none of the words considered by this procedure already disambiguated. We have to consider all the sense of both words in order to determine whether or not the distance between them is zero, this makes this procedure computationally intensive.
PROCEDURE 6: Form the semantic network [8] based on understanding made by the learning done from procedure #1 to procedure #5 and come to the final conclusion about the input sentence and action to be performed. The procedures presented above are applied iterative; this allows us to identify a set of nouns and verbs which can be disambiguated with high precision.
PROCEDURE 7: Using the procedures from PROCEDURE 1 to PROCEDURE 6, the system identifies the meaning of given input sentence. After identifying the actual meaning sentence (i.e. the action to be performed by the system) the system generates the SQL query for given input English text by correlating understudied meaning with actual database field which is exist in user requested database.
PROCEDURE 8: Using the procedures from PROCEDURE 1 to PROCEDURE 7, the system identifies the meaning of given input sentence and extracts the required data using SQL query from database. After the data extraction the query result data is projected in the report format as request in the user input English text.

The procedures presented above are applied iterative; this allows us to identify a set of nouns and verbs which can be disambiguated with high precision.

### 3.2 Object-Net Database Architecture

The existing knowledge bases in machine readable formats are WordNet, OMCSNet, MindNet, CYC, Thought treasure, VerbNet, Semcor, Open Mind Word Expert, Frame Net, and PropBank. These knowledge bases are useful to serve the purpose of developing information retrieval systems and shallow semantic representation for an input text. They model their elementary meanings only with conceptual world properties and constraints, and taxonomic relations between these words. They do not have synthesis capabilities, but rather their definitions are pre-programmed by humans. They do not make the machine creative enough to master its own language and to compose its own text based on its understood meanings. So a new methodology is required for machine to autonomously undertake the learning, analysis and of both the elementary and composite meanings of natural language, and most importantly, it is to note that the robustness of proposed algorithm by machine relies not only on sophisticated algorithms for knowledge manipulation but also the kind of knowledge it has. (i.e. careful modelling of elementary meanings from an engineering point of view). The new methodology for maintaining trained elementary meaning is called Object-Net database and details of this database is explained in analytical and synthesis capability section.

### 3.3 Algorithm with an Example

Consider for example to retrieve data from any of user database like "*I need the student report that joined on 04 November 2010 and report the result in the format of student name and course.*"

Procedure#1:
Tokenize the given sentence as below
"*I + need + the + student + report + that + joined + on + 04 + November + 2010 + and + report + the + result + in + the + format + of + student + name + and + course.*"
While categorizing these token words the below result is found.
"*Pro+Ver+Art+Nou+ver+pro+ver+adv+Num+Nov+Num+Con+Ver+Art+Ver+Pre+Art+Nov+Pre+Nov+Nov+Con+Nov*"
Creates the parse tree after tokenizing the sentence.

Procedure#2:
Find the words which are having unique sense of meaning and find object on which the action need to be performed.

"***I** (Sense#1) + **need** (Sense#1) + **the** (Sense#1) + student + report + **that** (Sense#1) + joined + (**on** + (**04** + **November** + **2010**)) (Sense#1) + and + report + the + result + **in** (Sense#1) + **the** (Sense#1) + **format** (Sense#1) + of + student + **name** (Sense#1) + and + course.*"
In this example the word "I", "the", "name", "that" and "date (04 November 2010)" are having only one sense of meaning, and student is the object on which the sentence related.

Procedur#3:

As per procedur#2 result, the related object or domain of sentence identified (i.e. as per example student), in Object-Net database search for the particular domain which is identified in procedur#2, from the identified object co-relate and identify meaning of the remaining words in sentence. Consider the network exist in object-net database as below fig 2.

While forming the two pairs one with the word previous to the current word and one next to the current word, for our example we will be arrived to the pairs as in below table 1, the last column shows that understanding.

Table 1: parsed tokens and its relation.

| S.No | Pairs | Description |
|---|---|---|
| 1 | I + need | Whom->I |
| 2 | need + the student | What -> the student |
| 3 | The student + report | What -> Report |
| 4 | Report + that | Unable to correlate |
| 5 | That + joined | Which-> joined |
| 6 | Joined + on | Which -> on |
| 7 | On + 04 November 2010 | Which -> date |
| 8 | And +report | What-> report |
| 9 | The + result | Which->result |
| 10 | Result + in | How -> in |
| 11 | In + the | Which -> the |
| 12 | The + format | What -> format |
| 13 | Format + of | How -> of |
| 14 | Of + student | What -> student |
| 15 | Student + name | What -> name |
| 16 | Name + and | Unable to correlate |
| 17 | And + course | What-> course |

Procedure#4:
From the procedure#3 we come to know that "need" is the action it required for "whom" is "I", "what" required is "student". From the student node "what" required is "report".
But "report" is ambiguous word in English it is having many meaning, and also by directly correlating words existing object-net is not giving correct path for the pairs "report + that" and "Name + and", as Date is already disambiguated and while considering pervious nodes it gives the meaning like "on" which is some date (ie. 04 November 2010).

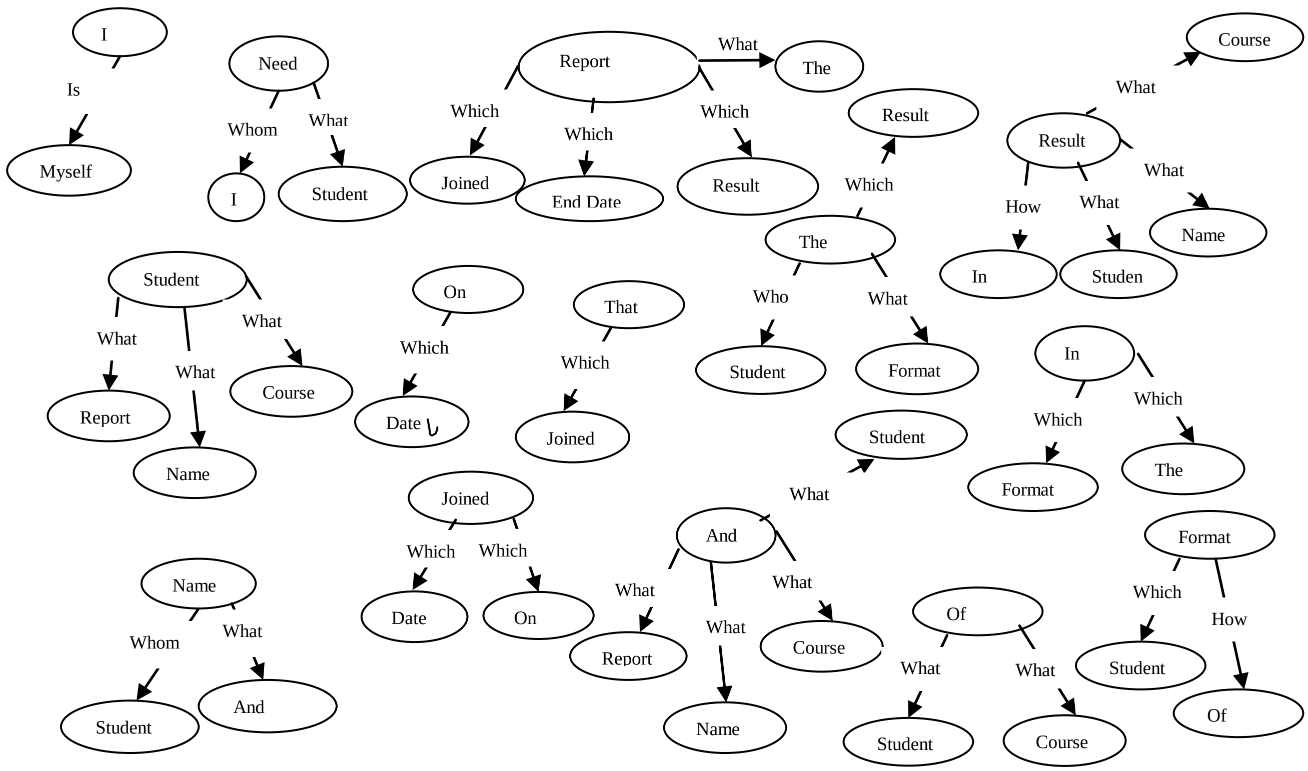


Fig. 2 Sample Object-Net Database

By node with connection distance of zero we will be arrived into the below mentioned paths.

1. I -> need
   a. I -> need -> the student
   b. I -> need-> the student -> report
2. That
   a. That -> joined
   b. That -> joined -> On
   c. That -> joined -> On -> 04 November 2010
   d. That -> joined -> On -> 04 November 2010 -> And
   e. That -> joined -> On -> 04 November 2010 -> And -> report
   f. That -> joined -> On -> 04 November 2010 -> And -> report -> the
   g. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result
   h. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in
   i. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the
   j. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format
   k. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format ->of
   l. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format ->of -> student
   m. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format ->of -> student -> name
3. And-> course

Procedur#5:
The word "report" was not clear still Procedur#4, now the report is clear like on "join date" some report is required. The ambiguous word "report" semantically connected with other part of the sentence in three ways as mentioned below.

1. Report -> joined -> 04 November 2010.
2. Report -> joined -> on -> 04 November 2010.
3. Report -> That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format ->of -> student -> name.

Here the path 2& 3 are already occurred in Procedur#4 but path 3 is bigger than path 2, so this path is considered, and now it is clear that report of joined date is required.

Procedur#6:
From the procedure#5, the "need" node is connected to "student" node. "student" node is connected to "report", "report" is connected to "joined date" and it is connected to "date", from this we have form a semantic network which gives the meaning as "need" is the action required by "I" and what required is "student", from "student" what required are report, and which report is "join date" report and similarly forms the semantic network.

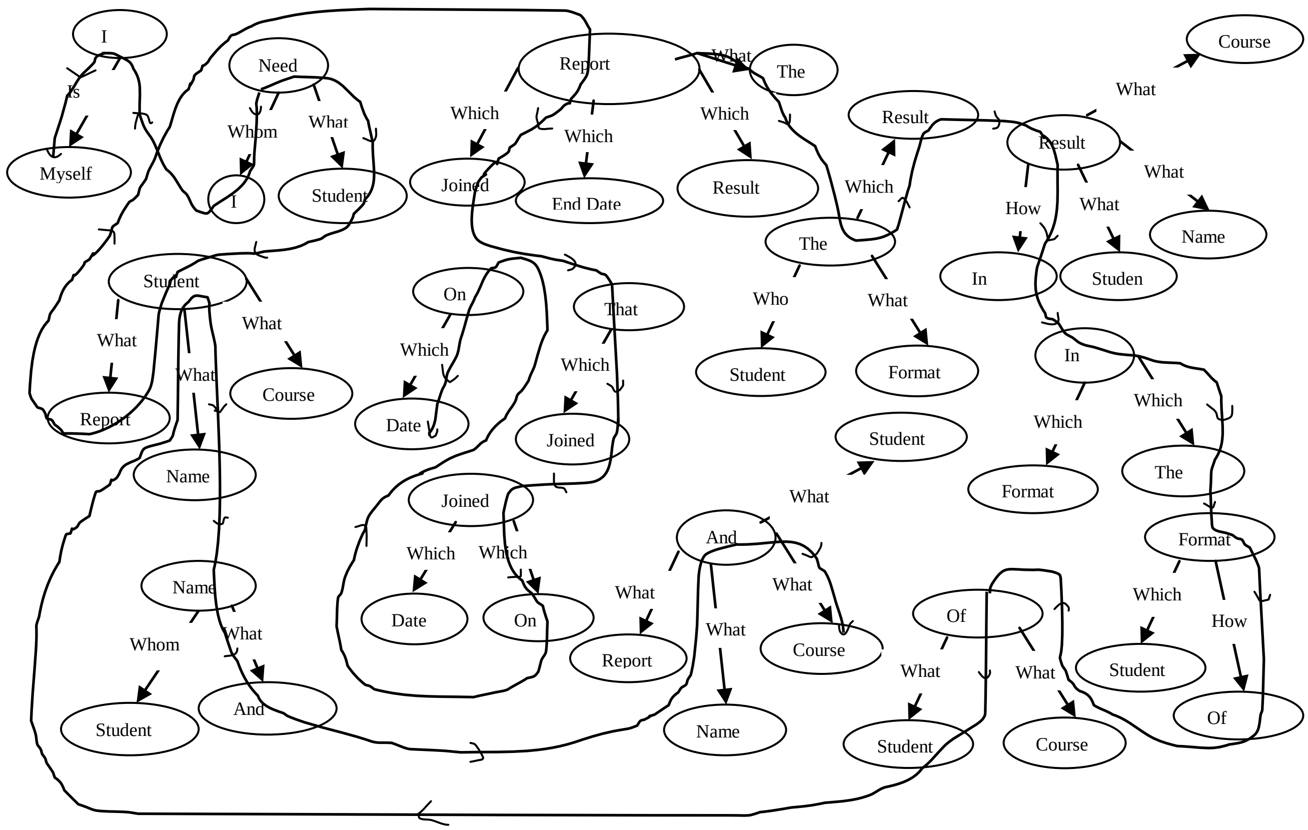


Fig. 3 Forming a semantic network in Object-Net Database

The above fig 3 shows the semantic relation path which gives meaning of the sentence.

Procedur#7:
After identifying the meaning and semantic relation path, system identifies that "Need" is the action and what need is "student" which is also exist in mapping list of user database, now identified the sub-user database as "Student details "but it contains two table as "Personal _details" and "mark details" at the same time from input text it identifies "joined date" which is existing in "Personal _details" so now it identifies table name "Personal _details" and writes the SQL statement as "From Personal _details" , the word "Report" and "student" says that selection of "name" and "course" is required so the system writes the SQL statement as "Select name, course", from the word "Report" it also identifies the SQL where class constraint as report of joined date is required but in input text contains some constant date value of "04 November 2010" so it writs the SQL statement as "Where Joined_date="04 November 2010"". While it recompiles all spices of

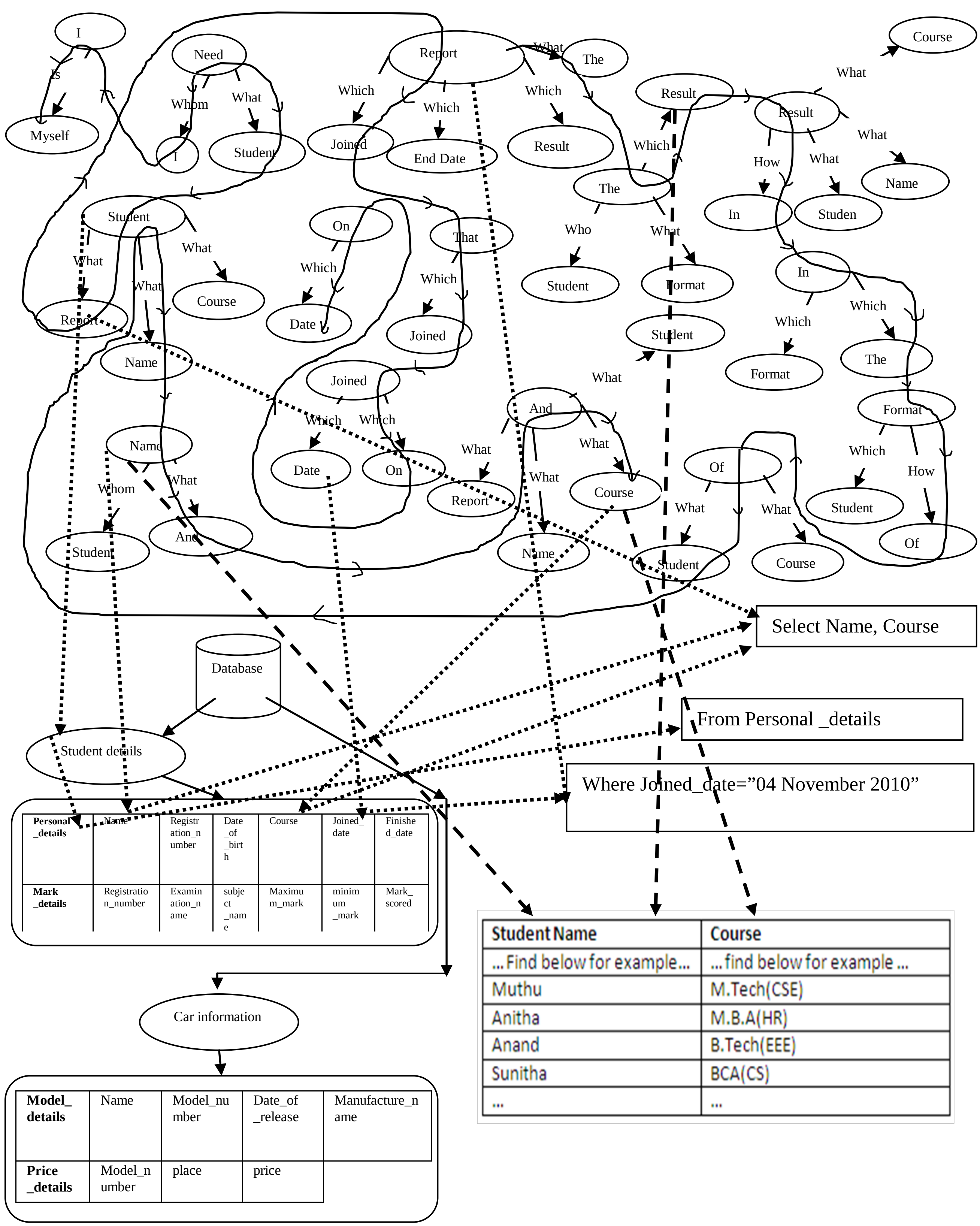


Fig. 4 Generating SQL statement from the semantic network in Object-Net Database

SQL statement which are generated as above procedures. The fig 4 shows the shows the semantic relation path along with SQL generation path. We get the complete SQL query as "*Select name, course From Personal _details Where Joined_date="04 November 2010"*. This SQL statement is executed by database engine where the user requested details are exist and gives result back to user.

Procedur#8:
The PROCEDURE 7 generates the SQL query based on the semantic network for reporting. From semantic network the system identifies the meaning of other part of the given sentence, from the "report" it understood that the "result" needs to be in "format" of "student name" and the "course". The result which came out after executing the SQL query is restricted to display only the student name and course as shown in the fig 4.

## 4. Analytical and Synthesis Capability In Object-Net Database

The example sentence "I need the student report that joined on 04 November 2010 and report the result in the format of student name and course" can be written in many as like mentioned below to reference same meaning as above sentence says. The possible ways are:

1. I need the student report that joined on 04 November 2010 and report the result in the format of student name and course
2. Need student report joined on 04 November 2010 and report the result in the format of student name and course
3. Report of student joined on 04 November 2010 in the format of student name and course
4. Student report joined on 04 November 2010 in the format of student name and course
5. On 04 November 2010 joined student report in the format of student name and course

The above mentioned sentences are giving same meaning as sentence#1, even though the sentences are not in corrected grammatical. But as a human can understand that meaning of all above sentence as "student name and course report is required who are all joined on 04 November 2010". So similarly we have to make sure that our proposed algorithm is also capable understanding the meaning of sentence as human.
For example the above sentence # 3 "Report of student joined on 04 November 2010 in the format of student name and course", in existing trained Object-Net network does not have direct relation from report and student but already the "what" relation were existing so it makes the new understanding link between "Report" and "student" with relation of "what". Similarly consider the above sentence#5 "On 04 November 2010 joined student report", this sentence starts with a date and it does not have action part like a action verb "need", in existing Object-Net doesn't have any of node starts with "Date" but there is a "Which relationship exists between "Joined" and "Date" so system creates a new node as "Date" to "Joined" with relation of "Which", next for student report there are two relationship exist one is from "Report" and another one from "student" node, now it creates two relation from newly created date "Date" node to "Student" and "Report" with relation of "Where" and "What" respectively. The fig 5 show the updated Object-Net database which will be used for future purpose.

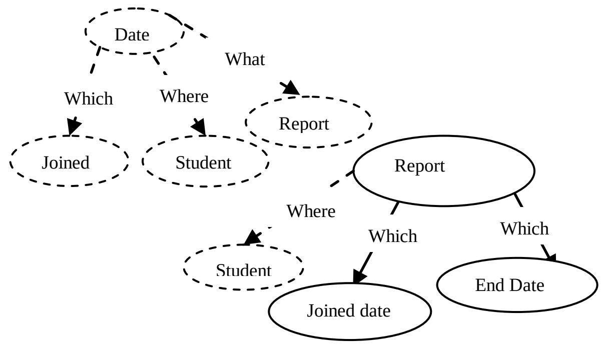


Fig. 5 updating active memory

So the system analyses and keeps updating its database memory there comes the system learning capability. If some words occurred in input text which is not exist in Object-Net database and also system is not able to resolve it internally then it will ask a master to train the relational network there come the human master into picture in order to correct and update the database.

## 5. Object-Net Approach for Database Extraction

We illustrate here the Object-Net disambiguation algorithm with the help of previous example "*I need the student report that joined on 04 November 2010 and report the result in the format of student name and course*". The system identifies the data meaning of the sentence, and what is the command, and what is action that user is expecting from the system. After identifying the meaning of the sentence, it maps the action to be done along with the trained internal actual database structure so that it can produce exact the SQL query for the input sentence or requirement. The bellow fig 6 shows that "student details" and "car information" databases are exist in a database; this mapping information is shared or trained to our system so that our system knows about where to fetch and which are to be fetched for a given sentence.

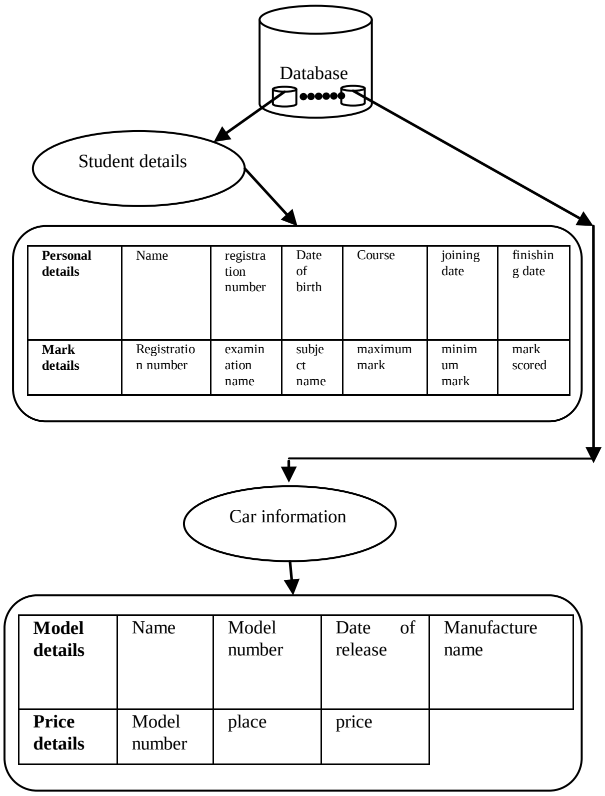


Fig. 6 Actual database information for mapping

As the system is capable of processing the English sentence, identify the meaning and writing the SQL query for the input sentence. In our above example instead of student details if the user asks for car details such as "I need the car details that are released on 01 MAY 2009" then the system will process this sentence and as above mentioned algorithm and it maps and writs SQL query on car details database. So the proposed system can be integrated to any database once the elementary meaning in Object-Net database is trained.

## 6. Performance of Data Extraction from Database, And Word Sense Disambiguation Based On Object-Net Database

The data extraction from database is the process of analyzing the requirement, and writing the SQL query based on the requirement. In our proposed algorithm the system itself analyzes the user requirement given interim of English sentence and writes the corresponding SQL queries. The performance of accuracy of writing SQL query and extracting the user required data is based on how the system understand the input text. But the system understanding of input sentence is based on the object-net database. The Object-net consists of set of initially trained entity network along with their meaningful representation with their action/behavior/property. The performance of our word sense disambiguation is mainly based on how many trained networks exist in Object-net database. If number of network data are high then number of hit ratio or number of occurrence of word in input text and trained network is high so it helps our algorithm to fetch correct object on which the input sentence is written and what is action or purpose of the sentence in order to give good accuracy on ambiguous words and sentence. When the number of trained network data of words in object-net database is less then number of hit ratio or number of occurrence of word in input text in trained network words is less so the active memory model of object-net database requires the help from master to train the non-trained words into database. So the accuracy the user requested data from user database is based on number trained network in Object-Net database and automatic system mapping between Object-Net database to actual system database entity fields. Fig 7 plots the graph between accuracy of the result of our algorithm versus number trained network word exist in object-net database, and the learning update required of object-net database in active memory model, along with accuracy over data extraction and automatic system mapping.

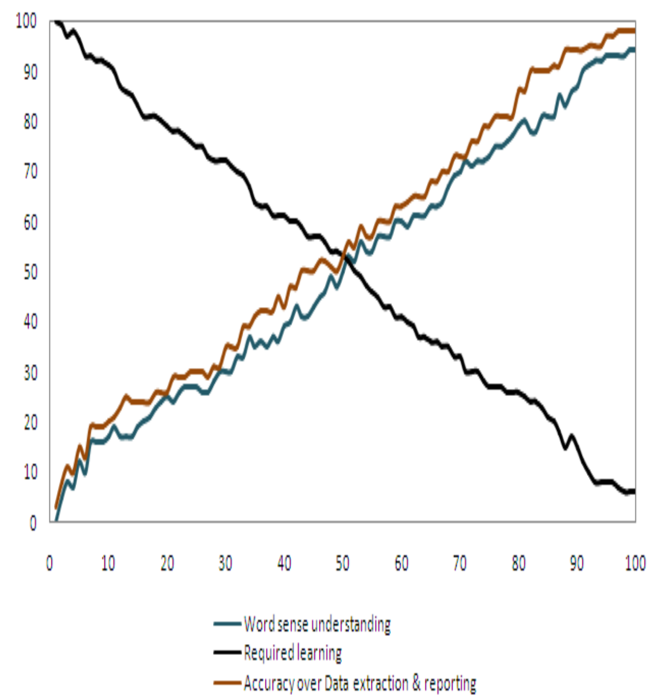


Fig. 7 Accuracy of Data extraction and Reporting Vs Number Trained Network, and automatic system mapping

## 7. Conclusion and Further Work

The algorithm identifies the meaning of requirement (i.e. user input sentence), analyzes the input sentence, and generates the SQL query statement, and reports the result as per the requested format. As per the above performance analyzes our algorithm extract and reports the data with precision of more than 97%, so it can be used in human computer interaction system, and data warehousing system in place of existing costliest commercial tools like DataStage, Business Object, etc.

## 8. Conclusion and Further Work

The algorithm identifies the meaning of sentence like human brain. It disambiguates ambiguous words based on object on which sentence is written and it generates the SQL query for data access and reporting. In future we can train our Object-net data base to other object or domains wherever intelligent human-computer interaction is required. And also from understanding of natural text meaning to the actual database query generation process can be implemented for accessing data from user database as per the user requirement.

## References

[1] Dan Klein, Kristina Toutanova, H. Tolga Ilhan, Sepandar D. Kamvar, and Christopher D.Manning, 2002 “Combining Heterogeneous Classifiers for Word-Sense Disambiguation.” In Workshop on Word Sense Disambiguation: Recent Successes and Future Directions at ACL 40, pages 74-80.

[2] Ezeife, C.I. and Y. Lu, 2005. Mining web log sequential patterns with position coded pre-order linked WAP-tree. Data Min. Knowl. Discov., 10: 5-38. DOI: 10.1007/s10618-005-0248-3

[3] Lee N, Laine A.F, Klein A, 2011.Towards a deep learning approach to brain parcellation, 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Page: 321 - 324, Chicago, IL, ISSN: 1945-7928.

[4] Sanderson, M., 1994. Word sense disambiguation and information retrieval. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 3-6, Springer-Verlag New York, Inc., New York, USA., pp: 142-151. http://portal.acm.org/citation.cfm?id=188548&dl=.

[5] Thomas, M. Connolly, Carolyn E. Begg and Carolyn E. Begg, 2004. Database Systems: A Practical Approach to Design, Implementation and Management. International Computer Science Ser.4th Edn., Pearson, ISBN-13: 9780321210258.

[6] Gordon S. Linoff and Michael J. A. Berry, 2011. Data Mining Techniques for Marketing, Sales and Customer Relationship Management. 3rd Edn., Wiley Computer Publishing, ISBN 0-470-65093-1.

[7] Jie Fan, Pu Wen, Hunan Radio & TV Univ., Changsha, 2007. Application of C4.5 Algorithm in Web-Based Learning Assessment System, IEEE International Conference on Machine Learning and Cybernetics, Issue Date: 19-22 Aug. 2007, On page(s): 4139 - 4143,Hong Kong, DOI: 10.1109/ICMLC.2007.4370871.

[8] Lee, G.N. and H. Fujita, 2007. K-means Clustering for Classifying Unlabelled MRI Data. Proceeding of the Digital Image Computing Techniques and Applications, Dec. 3-5, IEEE Xlpore Press, USA.,pp: 92-98. DOI: 10.1109/DICTA.2007.4426781

[9] Niknafs, A.A., M.E. Shiri and M.M. Javidi, 2005. An intelligent knowledge sharing strategy featuring item-based collaborative filtering and case based reasoning. Proceeding of the Intelligent Systems Design and Applications, Sept. 8-10, IEEE Xlpore Press, USA., pp: 67-72. DOI: 10.1109/ISDA.2005.22

[10] Claudio Giuliano, Alfio Massimiliano Gliozzo, and Carlo Strapparava, 2009. Kernel methods for minimally supervised wsd, ACM on Computational Linguistics, Volume 35 Issue 4, Pages 513-528,MIT Press Cambridge Publisher, MA, USA, ISSN: 0891-2017

[11] M. Barathi, S. Valli, 2010.Ontology Based Query Expansion Using Word Sense Disambiguation, International Journal of Computer Science and Information Security(IJCSIS) on Information Retrieval (cs.IR), Vol. 7, No. 2, pp. 022-027, USA, Computer Science ISSN 19475500.

**M.Munusamy**
He is a Ph.D research scholar in Department of Computer Science in Anna University- Coimbatore. His area of research is the application of natural language processing in database systems. He has M.Tech degree in Computer Science from SRM University- Chennai. Currently doing research on machine learning in computer science.

**G.Tholkappia Arasu**
He is having Ph.D on computer science. He is the principal of Jayam college of engineering and technology, Anna University, India. His area of research is the nanotechnology. He has several research publications in journals and presented papers in national and international conferences. Currently he is the chief editor of journal of convergence in engineering technology and science – JCETS.