# APPENDIX A

# Questionnaire

**QUESTIONNAIRE ON *DATABASE SYSTEMS***

**Introduction**

This questionnaire is part of an analysis of a research work in the area of Intelligent Tutoring Systems. The aim of the research is to improve students' performance and understanding in Database Systems. An intelligent tutoring systems is proposed which aims to improve the students skills in the area of Entity-Relationship Modelling (ER) and Normalization. This questionnaire aims to find out in general about the difficulty of the subject matters and the suitability of the proposed system.

**General information**

Please select your gender:

- ❑ Male
- ❑ Female

Please select your age group:

- ❑ 17-20
- ❑ 21-23
- ❑ 24-30
- ❑ above 30

At which level of studies do you learn the Database Systems Course?
*Please tick all that apply:*

- ❑ Diploma
- ❑ First Degree
- ❑ Master's

In your opinion, how important is a Database Systems course?
*Please rate in the scale below by ticking the appropriate box*:

Very important Not important

❑ ❑ ❑ ❑

**Experience with databases**:

How much experience do you have in using databases?

- ❑ 0-3 years
- ❑ 3-5 years
- ❑ more than 5 years

If any, what database systems have you used?
*Please tick all that apply:*

- ❑ Microsoft Access
- ❑ Visual Fox Pro

- ❑ INGRES
- ❑ ORACLE
- ❑ Others Please specify: _____

What type of query languages have you used?
*Please tick all that apply:*

- ❑ Structured Query Language (SQL)
- ❑ Query-By-Example (QBE)
- ❑ Quel
- ❑ XML query language
- ❑ ODMG OQL
- ❑ Others Please specify: _____

**Difficulty of subject in general:**

*Please rate in the scale below by ticking the appropriate box*:

(*Please ignore any topics which have not been covered during the course*)

| | Very difficult | | | Very easy |
|---|---|---|---|---|
| Introduction to Databases | ❑ | ❑ | ❑ | ❑ |
| Entity-Relationship Modelling | ❑ | ❑ | ❑ | ❑ |
| Normalization | ❑ | ❑ | ❑ | ❑ |
| The Relational Model | ❑ | ❑ | ❑ | ❑ |
| SQL | ❑ | ❑ | ❑ | ❑ |
| Relational Algebra and Calculus | ❑ | ❑ | ❑ | ❑ |
| File Organization and Storage Structures | ❑ | ❑ | ❑ | ❑ |

**Understanding according to topics in Entity-Relationship (ER) Modelling:**

Do you understand the basic concepts about entities, relationships and attributes?
*Please rate in the scale below by ticking the appropriate box*:

Very clear Very unclear

❑ ❑ ❑ ❑

How do you determine entities and relationships in ER modeling?

- ❑ Scanning the sentences for nouns, verbs etc.
- ❑ Visualizing the scenario
- ❑ Guessing
- ❑ Others Please specify: _____

Can you construct the ER model in a given problem/scenario?
*Please rate in the scale below by ticking the appropriate box*:

Most of the time Never

❑ ❑ ❑ ❑

**Understanding according to topics in Normalization:**

Do you understand the concept of functional dependency?
*Please rate in the scale below by ticking the appropriate box*:

Very clear Very unclear

❑ ❑ ❑ ❑

Do you know how to undertake the process of normalization?
*Please rate in the scale below by ticking the appropriate box*:

Most of the time Never

❑ ❑ ❑ ❑

**Use of diagramming tools:**

Have you used any diagramming tools to aid you in building an ER Model?

- ❑ Yes
- ❑ No

If yes, please specify: _____

Have you used any tools to aid you in carrying out normalization?

- ❑ Yes
- ❑ No

If yes, please specify: _____

Does the tools help in improving your skills in the subject matters?

- ❑ Yes
- ❑ No

**Problems encountered during the course:**

What sort of problems did you encounter during the course?
*Please tick all that apply:*

- ❑ Did not understand the lectures
- ❑ Lack of interest
- ❑ Feel bored with the subject
- ❑ Lack of skills
- ❑ None
- ❑ Others Please specify: _____

**Suitability of the system:**

Do you think an intelligent tutoring system that assists students in constructing ER models and normalizing relations would be useful?

- ❑ Yes
- ❑ No

*Thank you very much for your co-operation. Please give any comments about the questionnaire or any other information you wish to provide in the box provided below:*

Please return this questionnaire to:
Nazlia Omar
Room 2.4, 16 Malone Road, School of Computer Science, QUB

# APPENDIX B
# Penn Treebank II Part of Speech Tags

| | |
|---|---|
| CC | Coordinating Conjunction |
| CD | Cardinal number |
| DT | Determiner |
| EX | Existential "there" |
| FW | Foreign word |
| IN | Preposition or Subordinating conjunction |
| JJ | Adjective |
| JJR | Adjective, comparative |
| JJS | Adjective, superlative |
| LS | List item marker |
| MD | Modal |
| NN | Noun, singular or mass |
| NNS | Noun, plural |
| NNP | Proper noun, singular |
| NNPS | Proper noun, plural |
| PDT | Predeterminer |
| POS | Possessive ending |
| PRP | Personal pronoun |
| PRP$ | Possessive pronoun |
| RB | Adverb |
| RBR | Adverb, comparative |
| RBS | Adverb, superlative |
| RP | Particle |
| SYM | Symbol (mathematical or scientific) |
| TO | "to" |
| UH | Interjection |
| VB | Verb, base form |
| VBD | Verb, past tense |
| VBG | Verb, gerund or present participle |
| VBN | Verb, past participle |
| VBP | Verb, non 3rd-person singular present |
| VBZ | Verb, 3rd person singular present |
| WDT | Wh-determiner |
| WP | Wh-pronoun |
| WP$ | Possessive wh-pronoun |
| WRB | Wh-adverb |
| # | Pound sign |
| $ | Dollar sign |
| . | Sentence-final punctuation |
| , | Comma |
| : | Colon, semi-colon |
| ( | Left bracket character |
| ) | Right bracket character |
| " | Straight double quote |
| ` | Left open single quote |
| `` | Left open double quote |
| ' | Right close single quote |
| " | Right close double quote |

# Appendix C

# Training Dataset

## C.1 ARTICLES

An organization produces a number of journals. Each journal has a journal_name. Each journal may have any number of issues, and each issue is identified by its name and date_issued. Each issue contains a number of articles, which may be internally written or produced by external authors. A list of such external author names is kept and contains the author_name and author_ address. The fee_paid to an external author for an article is recorded, and the length in terms of number of words is kept for each article, together with the no_of_diagrams in the article. The department that produced an internal article is recorded. Each department has a department_name and a department_address.

## C.2 BUILDING

Each building in an organization has a different building_name and a building address. The meeting rooms in each building have their own room_no in the building, and each room has a specified seating_capacity. Rooms are available for hire for meetings, and each hire period must start on the hour. The hour and length_of_use are recorded. Each hire is made by a group in the organization, and groups are identified by a group_no and have a contact_phone. The facilities required for each hire period are also recorded. Each facility has an equip_no and a description.

## C.3 DEPARTMENT_PROJECT

Persons identified by a PERSON_ID and a SURNAME are assigned to departments identified by a DEPARTMENT_NAME. Persons work on projects, and each project has a PROJECT_ID and a BUDGET. Each project is managed by one department, and a department may manage many projects. But a person may work on only some (or none) of the projects in his or her department.

## C.4 DOCUMENT

Each person keeps a record of documents of interest. The time and source of each document are stored, along with its location. Documents may be books identified by

author and title or journal articles identified by journal volume and number, author and title or private correspondence identified by sender and date.

## C.5 HOSPITAL

A hospital wishes to computerize its information about staff, wards, patients and operations. Each patient has a unique number and the information such as his or her date of birth, address and occupation is stored. Each patient is allocated in a ward which has a unique number. The information about how many patients can be accommodated in a ward is recorded. Each patient could have more than one operation. The information recorded for each operation includes operation number, operation name, time, duration and the operation theatre number. A surgeon can perform many operations and a surgeon requires a number of nurses to assist. Each nurse also has a unique staff number and a number of nurses assigned can be assigned to a ward. For each member of staff, apart from the staff number, there is information such as date of birth, address, phone_no, salary etc. for him or her in the database.

## C.6 INSTRUCTOR_COURSE

Instructors may take a special loan of textbooks for courses that they teach. Each instructor has an instructor_name and a room_no, and each course has a course_name and a course_outline. Each textbook has a call_id and a title. More than one instructor can be assigned to a course, and more than one textbook can be used in each course. The books are assigned to the course, and an instructor always gets a special loan of all the books assigned to each course that he or she teaches.

## C.7 SUPPLIER

An organization purchases items from a number of suppliers. It keeps track of the number of each item type purchased from each supplier, and it also keeps a record of suppliers' addresses. Items are identified by item_type and have a description. There may be more than one such address for each supplier, and the price charged by each supplier for each item type is stored. Suppliers are identified by supplier_id.

## C.8 TRAINING COURSE

Each staff member attends a number of training courses. A training course is run by either one, two or three trainers and each trainer will run several different training courses. Each training course is held at just one of a number of possible venues and most venues hold several training courses. Every training course involves just one course theme and all course themes are involved in just one training course(with some themes recorded for some future allocation to an actual training course). Staff member details to be recorded include staff number, staff name, department, outcome of particular course and phone number. Trainer details include trainer code, trainer name, phone number and role on training course. Training course details include course code, course date and length. Course Theme includes theme number, theme name and theme area. Venue details include venue name, capacity and distance.

## C.9 VEHICLE_DRIVER

Each vehicle has a unique registration number and each driver a unique employee number. Drivers may be authorized to drive a number of vehicles, and any vehicle may be used by a number of drivers. Vehicles are allocated to departments within the company, although they may be used by drivers in other departments. Some classes of vehicle require specialist driver qualifications. There are occasional accidents which may lead to the vehicle being written off and/or the driver being disqualified from driving some or all classes of vehicle.

## C.10 VEHICLE_REGISTRATION

A person, identified by a person_id and a surname, can own any number of vehicles. Each vehicle is of a given MAKE and is registered in any one of a number of states identified by state_name. The registration number and the registration termination date are of interest, and so is the address of a registration office.

## C.11. WORKED EXAMPLE – Department project.doc

Persons identified by a PERSON_ID and a SURNAME are assigned to departments identified by a DEPARTMENT_NAME. Persons work on projects, and each project has a PROJECT_ID and a BUDGET. Each project is managed by one department, and a department may manage many projects. But a person may work on only some (or none) of the projects in his or her department.

| **Word** | **Shallow Parser (Output)** | **Meaning (tag)** | **Conversion of nouns into entities** | **Heuristics applied** | | **Result** |
|---|---|---|---|---|---|---|
| | | | | **Existing** | **New** | |
| Persons | Noun phrase | Proper singular noun | Entity | HE1 | | Entity |
| identified | Verb Phrase | Verb, past tense | | (Look for succeeding noun phrase-HA2) | | |
| by | Prepositional phrase | Preposition | | | | |
| a | Noun phrase | Determiner | | | | |
| PERSON_ID | | Proper singular noun | Entity | HA2 and HA1 | | Attribute (person) |
| and | | Coordinating conjunction | | | | |
| a | Noun phrase | Determiner | | | | |
| SURNAME | | Proper singular noun | Entity | HA2 | | Attribute (person) |
| are | Verb phrase | Verb, non 3rd ps. sing. present | | | | |
| assigned | | Verb, past participle | | | HR4 | assigned_to(?, department) |
| to | Prep. phrase | to | | | | |
| departments | NP | Plural noun | Entity | HE1 | | Entity |
| identified | VP | Verb, past participle | | | | |
| by | PP | Preposition | | | | |
| a | NP | Determiner | | | | |
| DEPARTMENT_NAME | | Proper singular noun | Entity | HA2 and HA1 | | Attribute (department) |
| . | | Final punctuation | | | | |
| Persons | ? NP | Proper singular noun | Entity | HE1 | | Entity (identified) |

| | | | | | | |
|---|---|---|---|---|---|---|
| work | VP | Verb, non 3rd ps. sing. present | | | HR4 | work_on (person,project) |
| on | PP | Preposition | | | | |
| projects | NP | Plural noun | Entity | HE1 | | Entity |
| , | | Comma | | | | |
| and | | Coordinating conjunction | | | | |
| each | NP | Determiner | | | | |
| project | | Singular noun | Entity | HE1 | | Entity (identified) |
| has | VP | Verb, 3rd ps. sing. present | | | | |
| a | NP | Determiner | | | | |
| PROJECT_ID | | Proper singular noun | Entity | HA1 and HA3 | | Attribute (project) |
| and | | Coordinating conjunction | | | | |
| a | NP | Determiner | | | | |
| BUDGET | | Proper singular noun | Entity | HA3 | | Attribute (project) |
| . | | Final punctuation | | | | |
| Each | NP | Determiner | | | | |
| project | | Singular noun | Entity | HE1 | | Entity (identified) |
| is | VP | Verb, 3rd ps. sing. present | | | | |
| managed | | Verb, past participle | | | HR4 | managed_by (project, department) |
| by | PP | Preposition | | | | |
| one | NP | Cardinal number | | | | |
| department | | Singular noun | Entity | HE1 | | Entity (identified) |
| , | | Comma | | | | |
| and | | Coordinating conjunction | | | | |
| a | NP | Determiner | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| department | | Singular noun | Entity | HE1 | | Entity (identified) |
| may | VP | Modal | | | | |
| manage | | Verb, base form | | | | |
| many | NP | Adjective | | | HC2 | Assign M to project |
| projects | | Plural noun | Entity | HE1 | | Entity (identified) |
| . | | Final punctuation | | | | |
| But | | Coordinating conjunction | | | | |
| a | NP | Determiner | | | | |
| person | | Singular noun | Entity | HE1 | | Entity (identified) |
| may | VP | Modal | | | | |
| work | | Verb, base form | | | HR4 | work_on(person,project) (identified) |
| on | PP | Preposition | | | | |
| only | NP | Adverb | | | | |
| some | | Determiner | | | | |
| or | | Coordinating conjunction | | | | |
| none | | Singular noun | Entity | | HEX | Non_entity |
| of | PP | Preposition | | | | |
| the | NP | Determiner | | | | |
| projects | | Plural noun | Entity | HE1 | | Entity (identified) |
| in | PP | Preposition | | | | |
| his | NP | Possesive pronoun | | | | |
| or | | Coordinating conjunction | | | | |
| her | | Possesive pronoun | | | | |
| department | | Singular noun | Entity | HE1 | | Entity (identified) |

| Word | Heuristics applied | Heuristics applied | Contribution | Contribution | Total weight | Result | Actual Answer | Correct | Incorrect | Ask user |
|---|---|---|---|---|---|---|---|---|---|---|
| Persons | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| identified by | | | | | | | | | | |
| a | | | | | | | | | | |
| PERSON_ID | HA2 | HA1 | New | New | -1.6 | Attribute | Attribute | 1 | 0 | 0 |
| and | | | | | | | | | | |
| a | | | | | | | | | | |
| SURNAME | HA2 | | New | | -0.7 | Attribute | Attribute | 1 | 0 | 0 |
| are | | | | | | | | | | |
| assigned to | HR4 | | Old | | 0.8 | Relationship | Relationship | 1 | 0 | 0 |
| departments | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| identified by | | | | | | | | | | |
| a | | | | | | | | | | |
| DEPARTMENT_NAME | HA2 | HA1 | New | New | -1.6 | Attribute | Attribute | 1 | 0 | 0 |
| . | | | | | | | | | | |
| Persons | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| work on | HR4 | | Old | | 0.8 | Relationship | Relationship | 1 | 0 | 0 |
| projects | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| , | | | | | | | | | | |
| and | | | | | | | | | | |
| each | | | | | | | | | | |
| project | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| has | | | | | | | | | | |
| a | | | | | | | | | | |
| PROJECT_ID | HA1 | HA3 | New | New | -1.8 | Attribute | Attribute | 1 | 0 | 0 |
| and | | | | | | | | | | |
| a | | | | | | | | | | |
| BUDGET | HA3 | | New | | -0.9 | Attribute | Attribute | 1 | 0 | 0 |
| . | | | | | | | | | | |
| Each | | | | | | | | | | |
| project | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| is | | | | | | | | | | |
| managed by | HR4 | | Old | | 0.8 | Relationship | Relationship | 1 | 0 | 0 |
| one | HC4 | | New | | 0.5 | Cardinality | Cardinality | 1 | 0 | 0 |
| department | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| , | | | | | | | | | | |
| and | | | | | | | | | | |
| a | | | | | | | | | | |
| department | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| may | | | | | | | | | | |
| manage | | | | | | | | | | |
| many | HC2 | | New | | 0.9 | Cardinality | Cardinality | 1 | 0 | 0 |
| projects | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| . | | | | | | | | | | |
| But | | | | | | | | | | |
| a | | | | | | | | | | |
| person | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| may | | | | | | | | | | |
| work on | HR4 | | Old | | 0.8 | Relationship | Relationship | 1 | 0 | 0 |
| only | | | | | | | | | | |
| some | | | | | | | | | | |
| or | | | | | | | | | | |
| none | HEX | | New | | 100 | Entity | Non entity | 0 | 1 | 0 |
| of | | | | | | | | | | |
| the | | | | | | | | | | |
| projects | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| in | | | | | | | | | | |
| his | | | | | | | | | | |
| or | | | | | | | | | | |
| her | | | | | | | | | | |
| department | HE1 | | Old | | 0.5 | Entity | Entity | 1 | 0 | 0 |
| | | **Total new** | **11** | **Total old** | **15** | **Total correct/ incorrect** | | **25** | **1** | **0** |

# Appendix D

# Test Dataset

## D.1. Airplane

### The problem:

Every airplane has a registration number and each airplane is of a specific model. The airport accommodates a number of airplane models. Each model is identified by a model number and has capacity and weight. A number of technicians work at the airport.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Airplane | Registration number |
| Airplane model | Model number, capacity, weight |
| Airport | |
| Technician | |
| *Relationship*<br>work at(technician, airport)<br>accommodates(airport, airplane model)<br>is_of(airplane, airplane model) | |
| *Cardinality* | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\_correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Airplane | 9 | 0 | 1 | 0 | 2 | 0 | 3 | 1 |

### Overall result:

Recall: 82%
Precision: 90%

### Source of error:

a) Undergenerated: missing relationship accomodates(airport, airplane models)
b) Undergenerated: missing relationship is_of(airplane, airplane model)
c) Incorrect: model identified as attribute
d) Unattached: attributes for airplane model(model number, capacity, weight)
e) Wrongly attached: attribute model attached to entity airplane

### Cases of multiple heuristics that are contradicting:

Nil

**The raw output from program:**

This is the output for file: airplane.txt

The entity is airplane/NN
The attributes are

registration/NN number/NN,model/NN,
airplane/NN has been applied with HE8.
at sentence 1.

It has the total weight of 0.7
The value is Entity
airplane/NN model/NN has been applied with HE7, HE7.
at sentence 2, 2.
It has the total weight of 1.2
The value is Entity
airport/NN has been applied with HE1.
at sentence 2.
It has the total weight of 0.5
The value is Entity
capacity/NN has been applied with HA2, HA3.
at sentence 3, 3.
It has the total weight of -1.6
The value is Attribute
model/NN has been applied with HA3, HE9.
at sentence 1, 3.
It has the total weight of -0.2
The value is Attribute
model/NN number/NN has been applied with HA8, HA2.
at sentence 3, 3.
It has the total weight of -1.5
The value is Attribute
number/NN has been applied with HEX, HEX.
at sentence 2, 4.
It has the total weight of 200
The value is Non entity
registration/NN number/NN has been applied with HA8, HA3.
at sentence 1, 1.
It has the total weight of -1.7
The value is Attribute
technicians/NNS has been applied with HE1.
at sentence 4.
It has the total weight of 0.5
The value is Entity
weight/NN has been applied with HA2, HA3.
at sentence 3, 3.
It has the total weight of -1.6
The value is Attribute
work/VBP at/IN has been applied with HR4.
at sentence 4.
It has the total weight of 0.8
The value is Relationship

From the relationship record:
The relationship is work/VBP at/IN
1st entity is technicians/NNS
2nd entity is airport/NN
at line 4

From the cardinality record:

## D.2. Bank

### The problem:

A bank planned to develop a database to maintain the following information about its customers, accounts, loans and mortgages. The bank is composed of branches identified by a unique code, a name and an address. The customers of the bank have accounts located at different branches. Each customer is unique identified by a code. The customer has a name, home phone number, work phone number, street address and zipcode. Each account belongs to only one customer.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Branches | Code, name, address |
| Customer | Code, name, home phone number, work phone number, street address, zipcode |
| Account | |
| Loan | |
| Mortgages | |
| *Relationship*<br>Belongs_to(account, customer) | |
| *Cardinality*<br>One(account, customer) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\_correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Bank | 14 | 1 | 2 | 1 | 0 | 1 | 0 | 0 |

### Overall result:

Recall: 88%
Precision: 74%

### Source of error:

a) Ask user: *account* has an initial value of Ask user
b) Incorrect: *bank*, the business environment is identified as an entity
c) Part_correct: *branches* identified as an attribute
d) Overgeneration of the entity *customer*
e) Incorrect: the relationship *planned to(bank, customer)*

### Cases of multiple heuristics that are contradicting:

a) Ask user: *account* has an initial value of Ask user

account/NN has been applied with HE1, HE1, HA3.
at sentence 6, 1, 3.
It has the total weight of 0.1

The value is Entity

**The raw output from program:**

This is the output for file: bank.txt

The entity is customer/NN
The attributes are
customer/NN code/NN,customer/NN name/NN,home/NN phone/NN number/NN,work/NN phone/NN number/NN,street/NN address/NN,zipcode/NN,

The entity is customers/NNS
The attributes are
branches/NNS,

account/NN has been applied with HE1, HE1, HA3.
at sentence 6, 1, 3.
It has the total weight of 0.1
The value is Entity
address/NN has been applied with HA2.
at sentence 2.
It has the total weight of -0.7
The value is Attribute
bank/NN has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
belongs/VBZ to/TO has been applied with HR4, HR5.
at sentence 6, 6.
It has the total weight of 1.6
The value is Relationship
branches/NNS has been applied with HE9, HA3.
at sentence 2, 3.
It has the total weight of -0.2
The value is Attribute
code/NN has been applied with HA2.
at sentence 2.
It has the total weight of -0.7
The value is Attribute
composed/VBN of/IN has been applied with HR4.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
customer/NN has been applied with HE9, HE8, HE1, HE8.
at sentence 4, 5, 1, 3.
It has the total weight of 2.6
The value is Entity
customer/NN code/NN has been applied with HA8, HA2.
at sentence 4, 4.
It has the total weight of -1.5
The value is Attribute
customer/NN name/NN has been applied with HA8, HA3.
at sentence 5, 5.
It has the total weight of -1.7
The value is Attribute
database/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
home/NN phone/NN number/NN has been applied with HA8, HA3.
at sentence 5, 5.
It has the total weight of -1.7
The value is Attribute
information/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
loans/NNS has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
located/VBN at/IN has been applied with HR4.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
mortgages/NNS has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
name/NN has been applied with HA2.
at sentence 2.
It has the total weight of -0.7
The value is Attribute
one/CD has been applied with HC4.
at sentence 6.
It has the total weight of 0.5
The value is Cardinality
planned/VBN to/TO has been applied with HR4.
at sentence 1.
It has the total weight of 0.8
The value is Relationship
street/NN address/NN has been applied with HA8, HA3.
at sentence 5, 5.
It has the total weight of -1.7
The value is Attribute

work/NN phone/NN number/NN has
been applied with HA8, HA3.
at sentence 5, 5.
It has the total weight of -1.7
The value is Attribute
zipcode/NN has been applied with HA3.
at sentence 5.
It has the total weight of -0.9
The value is Attribute

From the relationship record:
The relationship is planned/VBN to/TO
1st entity is bank/NN
2nd entity is customers/NNS
at line 1
The relationship is belongs/VBZ to/TO
1st entity is account/NN
2nd entity is customer/NN
at line 6

From the cardinality record:
The cardinality is one
(account/NN,customer/NN)
at line 6

## D.3. Boat hire

### The problem:

```
A boat-hire firm hires out boats of a range of types to customers. The
customer then takes a boat, which has a registration number, a name
and one numbered mooring on some river. A booking is associated with
one customer only.
```

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Boat | Registration number, name, mooring |
| Customer | |
| Booking | |
| *Relationship*<br>Associated with(booking, customer) | |
| *Cardinality*<br>One(booking, customer) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Boat_hire | 8 | 0 | 2 | 0 | 0 | 0 | 0 | 1 |

### Overall result:

Recall: 100%
Precision: 80%

### Source of error:

a) Incorrect: the business environment, *boat-hire firm* has been identified as an entity
b) Incorrect: *river* identified as an attribute
*c)* Wrongly attached entity, *customer*, to the attributes of *boat*

### Cases of multiple heuristics that are contradicting:

Nil

### The raw output from program:

```
This is the output for file: boat_hire.txt

The entity is customer/NN
The attributes are
registration/NN
number/NN,name/NN,mooring/NN,
river/NN,

associated/VBN with/IN has been applied
with HR4, HR5.
at sentence  3, 3.
```

It has the total weight of 1.6
The value is Relationship
boat/NN has been applied with HE1, HE1.
at sentence 2, 1.
It has the total weight of 1
The value is Entity
boat-hire/NN firm/NN has been applied with HE7.
at sentence 1.
It has the total weight of 0.6
The value is Entity
booking/NN has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
customer/NN has been applied with HE8, HE1.
at sentence 2, 1.
It has the total weight of 1.2
The value is Entity
has/VBZ has been applied with HR5.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
hires/VBZ out/RP has been applied with HR4.
at sentence 1.
It has the total weight of 0.8
The value is Relationship
mooring/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
name/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
one/CD has been applied with HC4, HC4.
at sentence 2, 3.
It has the total weight of 1
The value is Cardinality
registration/NN number/NN has been applied with HA8, HA3.
at sentence 2, 2.
It has the total weight of -1.7
The value is Attribute
river/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute

From the relationship record:
The relationship is associated/VBN
with/IN
1st entity is booking/NN
2nd entity is customer/NN
at line 3

From the cardinality record:
The cardinality is one
(booking/NN,customer/NN)
at line 3

## D.4. Bus

### The problem:

A country bus company owns a number of buses. Each bus is allocated to a particular route, although some route may have several buses. Each route passes through a number of towns. One or more drivers are allocated to each route.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Bus | |
| Route | |
| Driver | |
| Town | |
| *Relationship*<br>Allocated to(bus, route)<br>Allocated to(route, driver)<br>Passes through(route, town) | |
| *Cardinality*<br>Many(route, buses)<br>Many(driver, route) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Bus | 7 | 0 | 2 | 0 | 2 | 0 | 0 | 0 |

### Overall result:

Recall: 78%
Precision: 78%

### Source of error:

a) Incorrect: *buses* identified as attribute
b) Incorrect: the business environment, *country bus company* identified as entity
*c)* Undergenerated: the cardinality *many(route, buses)*
*d)* Undergenerated: the cardinality *many(driver, route)*

### Cases of multiple heuristics that are contradicting:

a) Buses incorrectly identified as an attribute:

buses/NNS has been applied with HE1, HA3.
at sentence 1, 2.
It has the total weight of -0.4
The value is Attribute

**The raw output from program:**

This is the output for file: bus.txt

allocated/VBN to/TO has been applied with HR4, HR4.
at sentence 2, 4.
It has the total weight of 1.6
The value is Relationship
bus/NN has been applied with HE8.
at sentence 2.
It has the total weight of 0.7
The value is Entity
buses/NNS has been applied with HE1, HA3.
at sentence 1, 2.
It has the total weight of -0.4
The value is Attribute
country/NN bus/NN company/NN has been applied with HE7.
at sentence 1.
It has the total weight of 0.6
The value is Entity
drivers/NNS has been applied with HE1.
at sentence 4.
It has the total weight of 0.5
The value is Entity
have/VB has been applied with HR5.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
number/NN has been applied with HEX, HEX.
at sentence 1, 3.
It has the total weight of 200
The value is Non entity
passes/VBZ through/IN has been applied with HR4.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
route/NN has been applied with HE1.
at sentence 2.
It has the total weight of 0.5
The value is Entity
several/JJ has been applied with HC2.
at sentence 2.
It has the total weight of 0.9
The value is Cardinality
towns/NNS has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity

From the relationship record:
The relationship is allocated/VBN to/TO
1st entity is bus/NN
2nd entity is route/NN
at line 2
The relationship is allocated/VBN to/TO
1st entity is drivers/NNS
2nd entity is route/NN
at line 4
The relationship is passes/VBZ through/IN
1st entity is route/NN
2nd entity is towns/NNS
at line 3

From the cardinality record:

## D.5. Cars

### The problem:

Each model is made up from many parts and each part may be used in the manufacture of more than one model. Each part has a description and an ID code. Each model of car is produced at just one of the firm's factories. A factory produces many models of car and many types of part although each type of part is produced at one factory only.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Model | |
| Part | description, ID code |
| Factory | |
| | |

*Relationship*
made up(model, parts)
produced at(model, factory)
produced at(part, factory)

*Cardinality*
Many(model,part)
Many(part, model)
Many(model,factory)
Many(part,factory)
one(factory,part)

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Cars | 13 | 0 | 6 | 1 | 0 | 0 | 0 | 1 |

### Overall result:

Recall: 100%
Precision: 65%

### Source of error:

a) Incorrect: *firm* identified as entity
b) Incorrect: *manufacture* identified as entity
c) Incorrect: the relationship *produced at(model, firm)*
d) Incorrect: the relationship *used in (part, manufacture)*
e) Incorrect: the cardinality *one (model, firm)*
f) Incorrect: the cardinality *one (manufacture, model)*
g) Overgenerated: the relationship *produces(factory, part)* ↔ *produced at (factory,part)*

h) Wrongly attached *manufacture* to cardinality many(part, model)

**Cases of multiple heuristics that are contradicting:**
Nil

**The raw output from program:**

This is the output for file: cars.txt

The entity is part/NN
The attributes are
description/NN,iD/NN code/NN,

car/NN has been applied with HE1, HE1, HE1, HE1, HE1, HE1.
at sentence 3, 4, 4, 3, 4, 4.
It has the total weight of 3
The value is Entity
description/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
factory/NN has been applied with HE1, HE1.
at sentence 4, 3.
It has the total weight of 1
The value is Entity
firm/NN has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
iD/NN code/NN has been applied with HA8, HA3.
at sentence 2, 2.
It has the total weight of -1.7
The value is Attribute
made/VBN up/RP has been applied with HR4, HR5.
at sentence 1, 1.
It has the total weight of 1.6
The value is Relationship
manufacture/NN has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
many/JJ has been applied with HC2, HC2.
at sentence 1, 4.
It has the total weight of 1.8
The value is Cardinality
model/NN has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
more/JJR than/IN has been applied with HC3.
at sentence 1.
It has the total weight of 0.6
The value is Cardinality
one/CD has been applied with HC4, HC4, HC4.
at sentence 1, 3, 4.
It has the total weight of 1.5
The value is Cardinality
part/NN has been applied with HE1, HE8, HE1, HE1, HE1, HE1, HE1, HE1, HE1.
at sentence 1, 2, 1, 4, 4, 4, 4, 4, 4.
It has the total weight of 4.7
The value is Entity
produced/VBN at/IN has been applied with HR4, HR5, HR4, HR5.
at sentence 3, 3, 4, 4.
It has the total weight of 3.2
The value is Relationship
produces/VBZ has been applied with HR5.
at sentence 4.
It has the total weight of 0.8
The value is Relationship
used/VBN in/IN has been applied with HR4.
at sentence 1.
It has the total weight of 0.8
The value is Relationship

From the relationship record:
The relationship is produced/VBN at/IN
1st entity is part/NN
2nd entity is factory/NN
at line 4
The relationship is made/VBN up/RP
1st entity is model/NN
2nd entity is parts/NNS
at line 1

The relationship is used/VBN in/IN
1st entity is part/NN
2nd entity is manufacture/NN
at line 1
The relationship is produces/VBZ
1st entity is factory/NN
2nd entity is part/NN
at line 4
The relationship is made/VBN up/RP
1st entity is model/NN
2nd entity is parts/NNS
at line 1
The relationship is produced/VBN at/IN

1st entity is model/NN
2nd entity is firm/NN
at line 3

From the cardinality record:
The cardinality is many
(factory/NN,part/NN)
at line 4
The cardinality is one
(model/NN,firm/NN)
at line 3
The cardinality is many
(model/NN,parts/NNS)
at line 1
The cardinality is one
(manufacture/NN,model/NN)
at line 1
The cardinality is one
(part/NN,factory/NN)
at line 4
The cardinality is many
(manufacture/NN,model/NN)
at line 1

## D.6. Client

### The problem:

```
An information service undertakes projects for clients. The client's
name and address are stored. A project has a project number, a project
name, deadline and a monetary value. It is led by a project leader who
supervises a team of employees.
```

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Project | Project number, name, deadline, monetary value |
| Employee | |
| Client | Name, address |
| *Relationship*<br>Managed by (project, employee) | |
| *Cardinality* | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Clientnew | 9 | 0 | 0 | 0 | 0 | 1 | 4 | 0 |

### Overall result:

Recall: 90%
Precision: 90%

### Source of error:

a) Ask user: *Client* has an initial value of *Ask User*
b) Unattached: the entities of the relationship *managed by(project, employee)*
c) Unattached: the attributes of *Client (name and address)*

### Cases of multiple heuristics that are contradicting:

a) Client has an initial value of 'Ask user'

client/NN has been applied with HA7, HE1.
at sentence 2, 1.
It has the total weight of 0
The value is Entity

**The raw output from program:**

This is the output for file: clientnew.txt

The entity is project/NN
The attributes are
project/NN number/NN,name/NN,deadline/NN,value/NN,

address/NN has been applied with HA7.
at sentence 2.
It has the total weight of -0.5
The value is Attribute
client/NN has been applied with HA7, HE1.
at sentence 2, 1.
It has the total weight of 0
The value is Entity
company/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
deadline/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9
The value is Attribute
managed/VBN by/IN has been applied with HR4.
at sentence 4.
It has the total weight of 0.8
The value is Relationship
name/NN has been applied with HA7, HA3.
at sentence 2, 3.
It has the total weight of -1.4
The value is Attribute
project/NN has been applied with HE8, HE1.
at sentence 3, 1.
It has the total weight of 1.2
The value is Entity
project/NN name/NN has been applied with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
project/NN number/NN has been applied with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
team/NN of/IN employees/NNS has been applied with HE1.
at sentence 4.
It has the total weight of 0.5
The value is Entity
value/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9
The value is Attribute

From the relationship record:

From the cardinality record:

## D.7. Company

### The problem:

A company has a number of departments. Each department consists of a number of employees, projects, and offices. Each employee has a job history. The employee also has a salary history. Each office has a number of phones. For each department, the department number, budget, and the manager's employee number are stored. For each employee, the employee number, current project number, office number, and phone number are stored. For each project, the project number and budget are stored. For each office, the office number, area in square feet, and all phone numbers are stored.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Department | Dept. number, budget, manager's employee number |
| Employee | Employee number, project number, office number, phone number, job history, salary history |
| Project | Project number, budget |
| Office | Office number, area, phone numbers |
| *Relationship* | |
| *Cardinality* | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Company | 16 | 1 | 0 | 0 | 0 | 1 | 2 | 0 |

### Overall result:

Recall: 89%
Precision: 89%

### Source of error:

a) Ask user: *project* has an initial value of Ask User
b) Incorrect: *department* identified as an attribute
c) Unattached: the attribute *budget* unattached to Project
d) Unattached: the attribute *area* unattached to Office

### Cases of multiple heuristics that are contradicting:

a) *Project* has an initial value of Ask user

project/NN has been applied with HA7, HE1.
at sentence 8, 2.

It has the total weight of 0
The value is Entity

b) *Job history* is correctly identified as an attribute:

job/NN history/NN has been applied with HE7, HA3.
at sentence 3, 3.
It has the total weight of -0.3
The value is Attribute

c) *Salary history* is correctly identified as an attribute:

salary/NN history/NN has been applied with HE7, HA3.
at sentence 4, 4.
It has the total weight of -0.3
The value is Attribute

**The raw output from program:**

This is the output for file: companynew.txt

The entity is employee/NN
The attributes are
job/NN history/NN,salary/NN history/NN,employee/NN number/NN,

The entity is project/NN
The attributes are
project/NN number/NN,

The entity is office/NN
The attributes are
phones/NNS,office/NN number/NN,

area/NN has been applied with HA7.
at sentence 9.
It has the total weight of -0.5
The value is Attribute
budget/NN has been applied with HA7, HA7.
at sentence 6, 8.
It has the total weight of -1
The value is Attribute
company/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
consists/VBZ of/IN has been applied with HR4.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
department/NN has been applied with HE1, HA7, HA3.
at sentence 2, 6, 1.
It has the total weight of -0.9
The value is Attribute
department/NN number/NN has been applied with HA8, HA7.
at sentence 6, 6.
It has the total weight of -1.3
The value is Attribute
employee/NN has been applied with HE8, HE8, HA7, HE1.
at sentence 3, 4, 7, 2.
It has the total weight of 1.4
The value is Entity
employee/NN number/NN has been applied with HA8, HA7, HA8, HA7.
at sentence 6, 6, 7, 7.
It has the total weight of -2.6
The value is Attribute
feet/NNS has been applied with HA7.
at sentence 9.
It has the total weight of -0.5
The value is Attribute
job/NN history/NN has been applied with HE7, HA3.
at sentence 3, 3.
It has the total weight of -0.3
The value is Attribute
manager/NN has been applied with HA7.
at sentence 6.
It has the total weight of -0.5
The value is Attribute
number/NN has been applied with HA3, HEX, HEX, HA3, HEX.
at sentence 1, 1, 2, 5, 5.
It has the total weight of 298.2
The value is Non entity
office/NN has been applied with HE8, HA7, HE1.
at sentence 5, 9, 2.
It has the total weight of 0.7
The value is Entity
office/NN number/NN has been applied with HA8, HA7, HA8, HA7.
at sentence 7, 7, 9, 9.
It has the total weight of -2.6
The value is Attribute
phone/NN number/NN has been applied with HA8, HA7, HA8, HA7, HA8, HA7.
at sentence 7, 7, 9, 9, 9, 9.

It has the total weight of -3.9
The value is Attribute
phones/NNS has been applied with HA3.
at sentence 5.
It has the total weight of -0.9
The value is Attribute
project/NN has been applied with HA7,
HE1.
at sentence 8, 2.
It has the total weight of 0
The value is Entity
project/NN number/NN has been applied
with HA8, HA7, HA8, HA7.
at sentence 7, 7, 8, 8.
It has the total weight of -2.6
The value is Attribute
salary/NN history/NN has been applied
with HE7, HA3.
at sentence 4, 4.
It has the total weight of -0.3
The value is Attribute

From the relationship record:

From the cardinality record:

## D.8. Computer

### The problem:

A company owns a number of branches which possess many computers to be rented. Customer makes booking for the computer of their choice. If a customer has made a booking, they are sent an invoice. An invoice may be associated with many payments. One payment may settle many invoices.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Branches | |
| Computer | |
| Customer | |
| Booking | |
| Invoice | |
| Payment | |
| *Relationship*<br>Possess(branch, computer)<br>Make(customer, booking)<br>Associated with(invoice, payment) | |
| *Cardinality*<br>Many(branches, computer)<br>Many(payment, invoices)<br>*Many(invoice, payment)* | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Computernew | 11 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |

### Overall result:

Recall: 92%
Precision: 85%

### Source of error:

a) Incorrect: *choice* identified as entity
b) Overgenerated: the relationship *settle(payment, invoices) ->associated with(invoice, payment)*
c) Undergenerated: relationship *make(customer, booking)*

### Cases of multiple heuristics that are contradicting:

Nil

### The raw output from program:

This is the output for file: computernew.txt

associated/VBN with/IN has been applied with HR4, HR5.
at sentence 4, 4.
It has the total weight of 1.6
The value is Relationship
booking/NN has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
booking/VBG for/IN has been applied with HR4.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
branches/NNS has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
choice/NN has been applied with HE1.
at sentence 2.
It has the total weight of 0.5
The value is Entity
company/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
computer/NN has been applied with HE1, HE1.
at sentence 2, 1.
It has the total weight of 1
The value is Entity
customer/NN has been applied with HE1, HE8.
at sentence 2, 3.
It has the total weight of 1.2
The value is Entity
invoice/NN has been applied with HE1, HE1.
at sentence 3, 5.
It has the total weight of 1
The value is Entity
many/JJ has been applied with HC2, HC2, HC2.
at sentence 1, 4, 5.
It has the total weight of 2.7
The value is Cardinality
number/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
payment/NN has been applied with HE1, HE1.
at sentence 5, 4.
It has the total weight of 1
The value is Entity
possess/VB has been applied with HR5.
at sentence 1.
It has the total weight of 0.8
The value is Relationship
settle/VB has been applied with HR5.
at sentence 5.
It has the total weight of 0.8
The value is Relationship

From the relationship record:
The relationship is associated/VBN with/IN
1st entity is invoice/NN
2nd entity is payments/NNS
at line 4
The relationship is possess/VB
1st entity is branches/NNS
2nd entity is computers/NNS
at line 1
The relationship is settle/VB
1st entity is payment/NN
2nd entity is invoices/NNS
at line 5

From the cardinality record:
The cardinality is many (branches/NNS,computers/NNS)
at line 1
The cardinality is many (payment/NN,invoices/NNS)
at line 5
The cardinality is many (invoice/NN,payments/NNS)
at line 4

## D.9. Doctor

### The problem:

```
Doctor prescribes drugs for patients. A doctor could prescribe one or
more drugs for several patients, and a patient could obtain
prescriptions from several doctors. Each prescription has a date and a
quantity associated with it.
```

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Doctor | |
| Prescription | date, quantity |
| Drug | |
| Patient | |
| *Relationship*<br>Prescribe(doctor, prescription) | |
| *Cardinality*<br>Many(doctor,patient)<br>Many(prescription, doctor) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Doctor | 8 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |

### Overall result:

Recall: 89%
Precision: 89%

### Source of error:

a) Incorrect: cardinality many(doctor,drugs )
b) Wrongly attached: entity *drugs* to the cardinality *many (doctor,patient*)
c) Undergenerated relationship *prescribe(doctor, prescription)*

### Cases of multiple heuristics that are contradicting:

Nil

### The raw output from program:

This is the output for file: doctor.txt

The entity is prescription/NN
The attributes are
date/NN,quantity/NN,

date/NN has been applied with HA3.
at sentence 4.
It has the total weight of -0.9
The value is Attribute
doctor/NN has been applied with HE1,
HE1,HE1.
at sentence 2, 3, 3.
It has the total weight of 1.5
The value is Entity
drugs/NNS has been applied with HE1.
at sentence 2.
It has the total weight of 0.5
The value is Entity
one/CD or/CC has been applied with HC4.
at sentence 3.
It has the total weight of 0.5
The value is Cardinality
patient/NN has been applied with HE1,
HE1.
at sentence 3, 2.
It has the total weight of 1
The value is Entity
prescription/NN has been applied with
HE8, HE1.
at sentence 4, 3.
It has the total weight of 1.2
The value is Entity
quantity/NN has been applied with HA3.
at sentence 4.
It has the total weight of -0.9
The value is Attribute
several/JJ has been applied with HC2.
at sentence 3.
It has the total weight of 0.9
The value is Cardinality

From the relationship record:

From the cardinality record:
The cardinality is many
(doctor/NN,drugs/NNS)
at line 3
The cardinality is many
(prescriptions/NNS,doctors/NNS)
at line 3
The cardinality is many
(drugs/NNS,patients/NNS)
at line 3

## D.10. Dreamhome

### The problem:

Each branch has a property for rent. The information stored on each property includes property number, address, type, monthly rent and property owner. Each property for rent is allocated to a member of staff, who oversees the management of the property. When a property is rented out, a rental agreement is drawn up between the renter and the property.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Property | Property number, address, type, monthly rent, property owner |
| Staff | |
| Rental agreement | |
| Renter | |
| *Relationship*<br>Allocated to (property, staff)<br>Drawn up (renter, property) | |
| *Cardinality* | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Dreamhome | 11 | 0 | 3 | 0 | 0 | 0 | 0 | 3 |

### Overall result:

Recall: 100%
Precision: 79%

### Source of error:

a) Incorrect: *branch* identified as entity
b) Incorrect: *management* identified as entity
c) Incorrect: relationship *rented out(property, agreement)*
d) Wrongly attached: attribute *rent* to entity *branch*
e) Wrongly attached: *agreement* attached to relationship *drawn up*
f) Wrongly attached: *management* attached to relationship *allocated to*

### Cases of multiple heuristics that are contradicting:

Nil

**The raw output from program:**

This is the output for file: dreamhome.txt

The entity is branch/NN
The attributes are
rent/NN,

The entity is property/NN
The attributes are
property/NN
number/NN,address/NN,type/NN,rent/
NN,property/NN owner/NN,
This is the output for file: dreamhome.txt

address/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
agreement/NN has been applied with HE1.
at sentence 4.
It has the total weight of 0.5
The value is Entity
allocated/VBN to/TO has been applied
with HR4.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
branch/NN has been applied with HE8.
at sentence 1.
It has the total weight of 0.7
The value is Entity
drawn/VBN up/RP has been applied with
HR4.
at sentence 4.
It has the total weight of 0.8
The value is Relationship
information/NN has been applied with
HEX.
at sentence 2.
It has the total weight of 100
The value is Non entity
management/NN has been applied with
HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
member/NN of/IN staff/NN has been
applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
property/NN has been applied with HE8.
at sentence 2.
It has the total weight of 0.7
The value is Entity
property/NN number/NN has been
applied with HA8, HA3.
at sentence 2, 2.
It has the total weight of -1.7
The value is Attribute
property/NN owner/NN has been applied
with HE7, HA3.
at sentence 2, 2.
It has the total weight of -0.3
The value is Attribute
rent/NN has been applied with HA3, HA3.
at sentence 1, 2.
It has the total weight of -1.8
The value is Attribute
rented/VBN out/RP has been applied with
HR4.
at sentence 4.
It has the total weight of 0.8
The value is Relationship
renter/NN has been applied with HE1.
at sentence 4.
It has the total weight of 0.5
The value is Entity
staff/NN has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
stored/VBN on/IN has been applied with
HR4.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
type/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute

From the relationship record:
The relationship is drawn/VBN up/RP
1st entity is agreement/NN
2nd entity is renter/NN
at line 4
The relationship is rented/VBN out/RP
1st entity is property/NN
2nd entity is agreement/NN
at line 4
The relationship is allocated/VBN to/TO
1st entity is property/NN
2nd entity is management/NN
at line 3

From the cardinality record:

From the plural record:

## D.11. Electronic supplier

### The problem:

Suppliers of electronic components supply many parts. However, each part is supplied by only one supplier. Parts are required for projects which have a project title and a deadline. Employees work on one project at a time. A part has a number and a description.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Supplier | |
| Part | Number, description |
| Project | Project title, deadline |
| Employee | |
| *Relationship*<br>Supplied by(supplier, part)<br>Required for (project, part)<br>Work on (employee, project) | |
| *Cardinality*<br>Many (supplier, part)<br>One (part, supplier)<br>One (employee, project) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\_correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Elect_ supplier | 14 | 0 | 3 | 0 | 0 | 0 | 4 | 2 |

### Overall result:

Recall: 100%
Precision: 82%

### Source of error:

a) Incorrect : *components* identified as entity
b) Incorrect: *time* identified as entity
c) Incorrect: the relationship *supply(components, parts)*
d) Unattached: the entities for the relationship *required for(project, part)*
e) Unattached: the entities for the relationship *work on(employee, project)*
f) Wrongly attached entity *Parts* instead of *Project* to the attributes *project title* and *deadline*
g) Wrongly attached *components* instead of *supplier* to the relationship *many (supplier, part)*

**Cases of multiple heuristics that are contradicting:**

a) *Project title* is correctly identified as an attribute

project/NN title/NN has been applied with HE7, HA3.
at sentence 3, 3.
It has the total weight of -0.3
The value is Attribute

**The raw output from program:**

This is the output for file:
electronic_supplier.txt

The entity is part/NN
The attributes are
part/NN number/NN,description/NN,

The entity is parts/NNS
The attributes are
project/NN title/NN,deadline/NN,

components/NNS has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
deadline/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9
The value is Attribute
description/NN has been applied with HA3.
at sentence 5.
It has the total weight of -0.9
The value is Attribute
employees/NNS has been applied with HE1.
at sentence 4.
It has the total weight of 0.5
The value is Entity
many/JJ has been applied with HC2.
at sentence 1.
It has the total weight of 0.9
The value is Cardinality
one/CD has been applied with HC4, HC4.
at sentence 2, 4.
It has the total weight of 1
The value is Cardinality
part/NN has been applied with HE1, HE8, HE1, HE8.
at sentence 2, 5, 1, 3.
It has the total weight of 2.4
The value is Entity
part/NN number/NN has been applied with HA8, HA3.
at sentence 5, 5.
It has the total weight of -1.7
The value is Attribute
project/NN has been applied with HE1, HE1.
at sentence 4, 3.
It has the total weight of 1
The value is Entity
project/NN title/NN has been applied with HE7, HA3.
at sentence 3, 3.
It has the total weight of -0.3
The value is Attribute
required/VBN for/IN has been applied with HR4.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
supplied/VBN by/IN has been applied with HR4, HR5.
at sentence 2, 2.
It has the total weight of 1.6
The value is Relationship
supplier/NN has been applied with HE1, HE1.
at sentence 2, 1.
It has the total weight of 1
The value is Entity
supply/VBP has been applied with HR5.
at sentence 1.
It has the total weight of 0.8
The value is Relationship
time/NN has been applied with HE1.
at sentence 4.
It has the total weight of 0.5
The value is Entity
work/VBP on/IN has been applied with HR4, HR5.
at sentence 4, 4.
It has the total weight of 1.6
The value is Relationship

From the relationship record:
The relationship is supply/VBP
1st entity is components/NNS
2nd entity is parts/NNS
at line 1
The relationship is supplied/VBN by/IN

1st entity is part/NN
2nd entity is supplier/NN
at line 2

From the cardinality record:

The cardinality is many
(components/NNS,parts/NNS)
at line 1
The cardinality is one
(part/NN,supplier/NN)
at line 2

## D.12. Employee

### The problem:

```
An employee is identified by an id. His name, address, telephone
number, job-title, date of joining and salary are to be kept. An
employee belongs to one or more departments. Each department has a
department name and a location.
```

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Employee | Id, name, address, telephone number, job title, date of joining, salary |
| Department | Name, location |
| *Relationship*<br>Belongs to (employee, department) | |
| *Cardinality*<br>Many (employee, department) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Employee | 12 | 0 | 0 | 0 | 1 | 0 | 6 | 0 |

### Overall result:

Recall: 92%
Precision: 100%

### Source of error:

a) Unattached: *name, address, telephone number, job title, date of joining* and *salary* are not attached to the entity *Employee*
b) Undergenerated: *job title* as an attribute

### Cases of multiple heuristics that are contradicting:

Nil

### The raw output from program:

This is the output for file: employee.txt

The entity is department/NN
The attributes are
name/NN,location/NN,

The entity is employee/NN
The attributes are
id/NN,

address/NN has been applied with HA7.
at sentence 2.
It has the total weight of -0.5
The value is Attribute
belongs/VBZ to/TO has been applied with HR4, HR5.
at sentence 3, 3.
It has the total weight of 1.6
The value is Relationship
date/NN has been applied with HA7.
at sentence 2.
It has the total weight of -0.5
The value is Attribute
department/NN has been applied with HE8, HE1.
at sentence 4, 3.
It has the total weight of 1.2
The value is Entity
department/NN name/NN has been applied with HA8, HA3.
at sentence 4, 4.
It has the total weight of -1.7
The value is Attribute
employee/NN has been applied with HE9.
at sentence 1.
It has the total weight of 0.7
The value is Entity
id/NN has been applied with HA2.
at sentence 1.
It has the total weight of -0.7
The value is Attribute
location/NN has been applied with HA3.
at sentence 4.
It has the total weight of -0.9
The value is Attribute
name/NN has been applied with HA7, HA3.
at sentence 2, 4.
It has the total weight of -1.4
The value is Attribute
one/CD or/CC has been applied with HC4.
at sentence 3.
It has the total weight of 0.5
The value is Cardinality
salary/NN has been applied with HA7.
at sentence 2.
It has the total weight of -0.5
The value is Attribute
telephone/NN number/NN has been applied with HA8, HA7.
at sentence 2, 2.
It has the total weight of -1.3
The value is Attribute

From the relationship record:
The relationship is belongs/VBZ to/TO
1st entity is employee/NN
2nd entity is departments/NNS
at line 3

From the cardinality record:
The cardinality is many
(employee/NN,departments/NNS)
at line 3

## D.13. Fault

### The problem:

An organization uses a number of equipment to produce goods. Each item is at one LOCATION, of one TYPE and has a DETAILED_DESCRIPTION. Faults on the equipment are identified by a unique FAULT_ID. Any number of persons may be assigned to a fault and work until it is fixed. Any number of parts may be used to repair a fault. The QTY-USED of each part is recorded against the fault. Each part is identified by a PART_ID, has a given WEIGHT and MAX-DIMENSION and can have any number of COLOURS. Each person is identified by a PERSON_ID, has a SURNAME and FIRST_NAME and any number of QUALIFICATIONS.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Item | Location, type, detailed description |
| Equipment | |
| Person | Person id, surname, first name, qualifications |
| Part | Part id, weight, dimension, colours, qty_used |
| Fault | Fault id |
| *Relationship*<br>Assigned to(person, fault)<br>Used to(part ,fault) | |
| *Cardinality*<br>Many (person, fault)<br>Many(part, fault) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Fault | 16 | 3 | 2 | 1 | 2 | 1 | 1 | 3 |

### Overall result:

Recall: 73%
Precision: 70%

### Source of error:

a) Part _correct: *detailed_description* identified as non entity
b) Part_correct: *location* identified as an entity
c) Part_correct: *type* identified as an entity
d) Incorrect: *goods* identified as an entity
e) Incorrect: *work* identified as an attribute
f) Undergenerated cardinality: *many(person, fault)*

g) Undergenerated cardinality: *many(part, fault)*
h) Overgenerated: relationship *recorded against(part, fault)*
i) Ask user: *fault* has an initial value of ask_user
j) Unattached: *qty_used* to entity *Part*
k) Wrongly attached: *fault id*, *fault* and *work* attached to *Equipment*

**Cases of multiple heuristics that are contradicting:**

a) DETAILED_DESCRIPTION has been incorrectly identified as a non-entity

dETAILED_DESCRIPTION/NNP has been applied with HA3, HEX.
at sentence 2, 2.
It has the total weight of 99.1
The value is Non entity

b) Fault has an initial value of ask_user

fault/NN has been applied with HA2, HE1.
at sentence 3, 3.
It has the total weight of -0.2
The value is Entity

**The raw output from program:**

This is the output for file: fault.txt

The entity is person/NN
The attributes are
pERSON_ID/NN,sURNAME/NNP,fIRST_
NAME/NNP,qUALIFICATIONS/NNP,

The entity is part/NN
The attributes are
pART_ID/NN,wEIGHT/NNP,mAX-
DIMENSION/NNP,cOLOURS/NNS,

The entity is equipment/NN
The attributes are
fAULT_ID/NN,work/NN,

any/DT has been applied with HC2, HC2.
at sentence 6, 7.
It has the total weight of 1.8
The value is Cardinality
assigned/VBN to/TO has been applied
with HR4.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
cOLOURS/NNS has been applied with
HA2, HA3.
at sentence 6, 6.
It has the total weight of -1.6
The value is Attribute
dETAILED_DESCRIPTION/NNP has been
applied with HA3, HEX.
at sentence 2, 2.
It has the total weight of 99.1
The value is Non entity
equipment/NN has been applied with
HE1, HE9.
at sentence 1, 3.
It has the total weight of 1.2
The value is Entity
fault/NN has been applied with HA2, HE1.
at sentence 3, 3.
It has the total weight of -0.2
The value is Entity
fAULT_ID/NN has been applied with
HA2.
at sentence 3.
It has the total weight of -0.7
The value is Attribute
fIRST_NAME/NNP has been applied with
HA2, HA3.
at sentence 7, 7.
It has the total weight of -1.6
The value is Attribute
given/VBN has been applied with HR5.
at sentence 6.
It has the total weight of 0.8
The value is Relationship
goods/NNS has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
item/NN has been applied with HE1.
at sentence 2.
It has the total weight of 0.5
The value is Entity

lOCATION/NNP has been applied with HE1.
at sentence 2.
It has the total weight of 0.5
The value is Entity
mAX-DIMENSION/NNP has been applied with HA2, HA3.
at sentence 6, 6.
It has the total weight of -1.6
The value is Attribute
number/NN has been applied with HEX, HA2, HEX, HEX, HA2, HA3, HEX, HA2, HA3, HEX.
at sentence 1, 3, 3, 4, 6, 6, 6, 7, 7, 7.
It has the total weight of 496.1
The value is Non entity
one/CD has been applied with HC4.
at sentence 2.
It has the total weight of 0.5
The value is Cardinality
organization/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
part/NN has been applied with HE1, HE9, HE8, HE1.
at sentence 5, 6, 6, 4.
It has the total weight of 2.4
The value is Entity
pART_ID/NN has been applied with HA2.
at sentence 6.
It has the total weight of -0.7
The value is Attribute
person/NN has been applied with HE9, HE8, HA2.
at sentence 7, 7, 3.
It has the total weight of 0.7
The value is Entity
pERSON_ID/NN has been applied with HA2.
at sentence 7.
It has the total weight of -0.7
The value is Attribute
qTY-USED/NNP has been applied with HA7.
at sentence 5.
It has the total weight of -0.5
The value is Attribute
qUALIFICATIONS/NNP has been applied with HA2, HA3.
at sentence 7, 7.
It has the total weight of -1.6
The value is Attribute
recorded/VBN against/IN has been applied with HR4.
at sentence 5.
It has the total weight of 0.8
The value is Relationship
sURNAME/NNP has been applied with HA2, HA3.
at sentence 7, 7.
It has the total weight of -1.6
The value is Attribute
tYPE/NN has been applied with HE1.
at sentence 2.
It has the total weight of 0.5
The value is Entity
used/VBN to/TO has been applied with HR4.
at sentence 4.
It has the total weight of 0.8
The value is Relationship
wEIGHT/NNP has been applied with HA2, HA3.
at sentence 6, 6.
It has the total weight of -1.6
The value is Attribute
work/NN has been applied with HA2.
at sentence 3.
It has the total weight of -0.7
The value is Attribute

From the relationship record:
The relationship is assigned/VBN to/TO
1st entity is persons/NNS
2nd entity is fault/NN
at line 3
The relationship is recorded/VBN against/IN
1st entity is part/NN
2nd entity is fault/NN
at line 5
The relationship is used/VBN to/TO
1st entity is parts/NNS
2nd entity is fault/NN
at line 4

From the cardinality record:

## D.14. Hospital

### The problem:

A hospital is organized into a number of wards, each containing many patients. Each ward is staffed by nurses. Each doctor belongs to a group. Each doctor is responsible for many patients. A doctor may treat patients from many wards.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Ward | |
| Nurse | |
| Patient | |
| Doctor | |
| Group | |
| *Relationship*<br>Staffed by (ward, nurse)<br>Responsible for(doctor, patient)<br>Contain(ward, patient)<br>Belongs to(doctor, group) | |
| *Cardinality*<br>Many(doctor, patient)<br>Many(ward, patient) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Hospitalnew | 10 | 0 | 3 | 0 | 1 | 0 | 0 | 0 |

### Overall result:

Recall: 91%
Precision: 77%

### Source of error:

a) Incorrect: *hospita*l, the business environment, identified as an entity
b) Incorrect: the relationship *organized into(hospital, wards)*
c) Incorrect: the cardinality *many(patients, wards)*
d) Undergenerated: the relationship *responsible for(doctor, patient)*

### Cases of multiple heuristics that are contradicting:

Nil

### The raw output from program:

This is the output for file: hospitalnew.txt

belongs/VBZ to/TO has been applied with HR4.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
containing/VBG has been applied with HR5.
at sentence 1.
It has the total weight of 0.8
The value is Relationship
doctor/NN has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
group/NN has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
hospital/NN has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
many/JJ has been applied with HC2, HC2, HC2.
at sentence 1, 4, 5.
It has the total weight of 2.7
The value is Cardinality
number/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
nurses/NNS has been applied with HE1.
at sentence 2.
It has the total weight of 0.5
The value is Entity
organized/VBN into/IN has been applied with HR4.
at sentence 1.
It has the total weight of 0.8
The value is Relationship
patients/NNS has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
staffed/VBN by/IN has been applied with HR4.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
ward/NN has been applied with HE1, HE1.
at sentence 2, 1.
It has the total weight of 1
The value is Entity

From the relationship record:
The relationship is belongs/VBZ to/TO
1st entity is doctor/NN
2nd entity is group/NN
at line 3
The relationship is organized/VBN into/IN
1st entity is hospital/NN
2nd entity is wards/NNS
at line 1
The relationship is staffed/VBN by/IN
1st entity is ward/NN
2nd entity is nurses/NNS
at line 2
The relationship is containing/VBG
1st entity is wards/NNS
2nd entity is patients/NNS
at line 1

From the cardinality record:
The cardinality is many
(wards/NNS,patients/NNS)
at line 1
The cardinality is many
(patients/NNS,wards/NNS)
at line 5
The cardinality is many
(doctor/NN,patients/NNS)
at line 4

## D.15. Invoice

### The problem:

An invoice is written by a sales representative. Each sales representative can write many invoices, but each invoice is written by a single sales representative. The invoice is written for a single customer. However, each customer may have many invoices. An invoice may contain many products. Each product can be associated with many vendors.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Invoice | |
| Sales representative | |
| Customer | |
| Product | |
| Vendor | |

*Relationship*
Written by (invoice, sales representative)
Written for (invoice, customer)
Contain (invoice, product)
Associated with (product, vendor)
Receive (customer, invoice)

*Cardinality*
many (sales representative, invoice)
one (invoice, sales representative)
one (invoice, customer)
many (customer, invoice)
many (invoice, product)
many (product, vendor)

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Invoice | 14 | 0 | 0 | 0 | 2 | 0 | 2 | 0 |

### Overall result:

Recall: 88%
Precision: 100%

### Source of error:

a) Undergenerated cardinality: *many(sales representative, invoice)*
b) Undergenerated cardinality: *one(invoice, sales representative)*
c) Unattached entities to relationship *written by(invoice, sales representative)*

**Cases of multiple heuristics that are contradicting:**
Nil

**The raw output from program:**
This is the output for file: invoice.txt

associated/VBN with/IN has been applied with HR4, HR5.
at sentence 6, 6.
It has the total weight of 1.6
The value is Relationship
contain/VB has been applied with HR5.
at sentence 5.
It has the total weight of 0.8
The value is Relationship
customer/NN has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
invoice/NN has been applied with HE1, HE1.
at sentence 1, 2.
It has the total weight of 1
The value is Entity
many/JJ has been applied with HC2, HC2, HC2, HC2.
at sentence 2, 4, 5, 6.
It has the total weight of 3.6
The value is Cardinality
product/NN has been applied with HE1, HE1.
at sentence 6, 5.
It has the total weight of 1
The value is Entity
receive/VB has been applied with HR5.
at sentence 4.
It has the total weight of 0.8
The value is Relationship
sales/NNS representative/NN has been applied with HE7, HE7.
at sentence 1, 2.
It has the total weight of 1.2
The value is Entity
single/JJ has been applied with HC4, HC4.
at sentence 2, 3.
It has the total weight of 1
The value is Cardinality
vendors/NNS has been applied with HE1.
at sentence 6.
It has the total weight of 0.5
The value is Entity
written/VBN by/IN has been applied with HR4, HR4, HR5.
at sentence 1, 2, 2.
It has the total weight of 2.4
The value is Relationship
written/VBN for/IN has been applied with HR4, HR5.
at sentence 3, 3.
It has the total weight of 1.6
The value is Relationship

From the relationship record:
The relationship is associated/VBN
with/IN
1st entity is product/NN
2nd entity is vendors/NNS
at line 6
The relationship is receive/VB
1st entity is customer/NN
2nd entity is invoices/NNS
at line 4
The relationship is written/VBN for/IN
1st entity is invoice/NN
2nd entity is customer/NN
at line 3
The relationship is contain/VB
1st entity is invoice/NN
2nd entity is products/NNS
at line 5

From the cardinality record:
The cardinality is many
(invoice/NN,products/NNS)
at line 5
The cardinality is many
(customer/NN,invoices/NNS)
at line 4
The cardinality is one
(invoice/NN,customer/NN)
at line 3
The cardinality is many
(product/NN,vendors/NNS)
at line 6

## D.16. Library

### The problem:

A library has many book suppliers. Each book suppliers supplies many books. A book may be obtained from one or more book suppliers. A book supplier may receive many orders. A book has one ISBN, a title and a single publisher. The supplier's name and address are stored. The supplier is identified by a supplier number. The order has an order number.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Book supplier | Name, address, supplier number |
| Book | ISBN, title, publisher |
| Order | Order number |
| *Relationship*<br>supplies (book supplier, supply)<br>receive(book supplier, order) | |
| *Cardinality*<br>many (book supplier, book)<br>many (book, book supplier)<br>many(book supplier, order) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Library | 14 | 0 | 3 | 0 | 1 | 0 | 2 | 3 |

### Overall result:

Recall: 93%
Precision: 82%

### Source of error:

a) Incorrect: the business environment*, library*, has been identified as an entity
b) Incorrect cardinality: many (library, book)
c) Incorrect relationship: has (library, book)
d) Undergenerated cardinality many (book supplier, book)
e) Unattached: name and address are not attached to book supplier
f) Wrongly attached: book instead of book supplier in cardinality many (book supplier, order)
g) Wrongly attached: book instead of book supplier in cardinality many (book supplier, book)
h) Wrongly attached book in relationship supplies (book, books)- should be book supplier

i) Wrongly attached book in relationship receive (book, orders)- should be book supplier

**Cases of multiple heuristics that are contradicting:**

a) Book supplier is correctly identified as an entity

book/NN supplier/NN has been applied with HE7, HE7, HA7, HE7, HE9, HE7, HA3, HE7, HE7, HE7, HA3, HE7, HE7.
at sentence 4, 6, 6, 7, 7, 1, 1, 2, 3, 1, 1, 2, 3.
It has the total weight of 3.8
The value is Entity

**The raw output from program:**

This is the output for file: library.txt

The entity is order/NN
The attributes are
order/NN number/NN,

The entity is book/NN
The attributes are
iSBN/NN,title/NN,publisher/NN,

The entity is book/NN supplier/NN
The attributes are
supplier/NN number/NN,

address/NN has been applied with HA7.
at sentence 6.
It has the total weight of -0.5
The value is Attribute
book/NN has been applied with HE1, HE8, HE1.
at sentence 3, 5, 2.
It has the total weight of 1.7
The value is Entity
book/NN supplier/NN has been applied with HE7, HE7, HA7, HE7, HE9, HE7, HA3, HE7, HE7, HE7, HA3, HE7, HE7.
at sentence 4, 6, 6, 7, 7, 1, 1, 2, 3, 1, 1, 2, 3.
It has the total weight of 3.8
The value is Entity
has/VBZ has been applied with HR5, HR5.
at sentence 1, 5.
It has the total weight of 1.6
The value is Relationship
iSBN/NN has been applied with HA3.
at sentence 5.
It has the total weight of -0.9
The value is Attribute
library/NN has been applied with HE8.
at sentence 1.
It has the total weight of 0.7
The value is Entity
many/JJ has been applied with HC2, HC2, HC2.
at sentence 1, 2, 4.
It has the total weight of 2.7
The value is Cardinality
name/NN has been applied with HA7.
at sentence 6.
It has the total weight of -0.5
The value is Attribute
obtained/VBN from/IN has been applied with HR4, HR5.
at sentence 3, 3.
It has the total weight of 1.6
The value is Relationship
one/CD has been applied with HC4.
at sentence 5.
It has the total weight of 0.5
The value is Cardinality
one/CD or/CC has been applied with HC4.
at sentence 3.
It has the total weight of 0.5
The value is Cardinality
order/NN has been applied with HE8, HE1.
at sentence 8, 4.
It has the total weight of 1.2
The value is Entity
order/NN number/NN has been applied with HA8, HA3.
at sentence 8, 8.
It has the total weight of -1.7
The value is Attribute
publisher/NN has been applied with HA3.
at sentence 5.
It has the total weight of -0.9
The value is Attribute
receive/VB has been applied with HR5.
at sentence 4.
It has the total weight of 0.8
The value is Relationship
single/JJ has been applied with HC4.
at sentence 5.
It has the total weight of 0.5
The value is Cardinality

supplier/NN number/NN has been
applied with HA8, HA2.
at sentence 7, 7.
It has the total weight of -1.5
The value is Attribute
supplies/VB has been applied with HR5.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
title/NN has been applied with HA3.
at sentence 5.
It has the total weight of -0.9
The value is Attribute

From the relationship record:
The relationship is has/VBZ
1st entity is library/NN
2nd entity is book/NN
at line 1
The relationship is supplies/VB
1st entity is book/NN
2nd entity is books/NNS
at line 2
The relationship is receive/VB
1st entity is book/NN
2nd entity is orders/NNS
at line 4

From the cardinality record:
The cardinality is many
(book/NN,orders/NNS)
at line 4
The cardinality is many
(library/NN,book/NN)
at line 1

## D.17. Library books

### The problem:

Each library book has a unique code number and is specified in a particular category. The information stored for each library book includes its title, author, publisher and year of publication. Each borrower has a unique borrower number together with his/her name, address and occupation. Each library is identified by its name, but information such as its location and quantity of books in it is also stored. Each library can order books from a number of publishers and each publisher supplies books for a number of libraries. Each publisher has a unique name, location and year of establishment to be stored.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Library book | Code number, category, title, author, publisher, year of publication |
| Borrower | Borrower number, name, address, occupation |
| Library | Name, location, quantity of books |
| Publishers | Name, location, year of establishment |
| *Relationship* | |
| *Cardinality*<br>Many(library, publisher)<br>Many(publisher,library) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Librarybook | 20 | 0 | 1 | 0 | 2 | 0 | 1 | 0 |

### Overall result:

Recall: 91%
Precision: 95%

### Source of error:

a) Incorrect: the attribute *book*
b) Undergenerated: cardinality many(library, publisher)
c) Undergenerated: cardinality many(publisher, library)
d) Unattached: *publisher* to the entity *library book*

### Cases of multiple heuristics that are contradicting:

a) Information is correctly identified as a non-entity

information/NN has been applied with HEX, HA2, HA7, HEX.
at sentence 2, 4, 4, 4.
It has the total weight of 198.8
The value is Non entity

b) Library is correctly identified as an entity

library/NN has been applied with HE9, HA7, HE1.
at sentence 4, 4, 5.
It has the total weight of 0.7
The value is Entity

c) Publisher is correctly identified as an attribute

publisher/NN location/NN has been applied with HE7, HA3.
at sentence 6, 6.
It has the total weight of -0.3
The value is Attribute

d) Book is incorrectly identified as an attribute

book/NN has been applied with HE1, HA2, HA7.
at sentence 5, 4, 4.
It has the total weight of -0.7
The value is Attribute

**The raw output from program:**

This is the output for file: librarybooks.txt

The entity is library/NN
The attributes are
name/NN,location/NN,quantity/NN
of/IN books/NNS,

The entity is publisher/NN
The attributes are
publisher/NN name/NN,publisher/NN
location/NN,year/NN of/IN
establishment/NN,

The entity is library/NN book/NN
The attributes are
code/NN
number/NN,category/NN,title/NN,
author/NN,year/NN of/IN
publication/NN,

The entity is borrower/NN
The attributes are
borrower/NN number/NN,borrower/NN
name/NN,address/NN,occupation/NN,

address/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9

The value is Attribute
author/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
book/NN has been applied with HE1,
HA2, HA7.
at sentence 5, 4, 4.
It has the total weight of -0.7
The value is Attribute
borrower/NN has been applied with HE8.
at sentence 3.
It has the total weight of 0.7
The value is Entity
borrower/NN name/NN has been applied
with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
borrower/NN number/NN has been
applied with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
category/NN has been applied with HA3.
at sentence 1.

It has the total weight of -0.9
The value is Attribute
code/NN number/NN has been applied with HA8, HA3.
at sentence 1, 1.
It has the total weight of -1.7
The value is Attribute
information/NN has been applied with HEX, HA2, HA7, HEX.
at sentence 2, 4, 4, 4.
It has the total weight of 198.8
The value is Non entity
library/NN has been applied with HE9, HA7, HE1.
at sentence 4, 4, 5.
It has the total weight of 0.7
The value is Entity
library/NN book/NN has been applied with HE7, HE8, HE7, HE8.
at sentence 1, 1, 2, 2.
It has the total weight of 2.6
The value is Entity
location/NN has been applied with HA2, HA7.
at sentence 4, 4.
It has the total weight of -1.2
The value is Attribute
name/NN has been applied with HA2, HA7.
at sentence 4, 4.
It has the total weight of -1.2
The value is Attribute
number/NN has been applied with HEX.
at sentence 5.
It has the total weight of 100
The value is Non entity
occupation/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9
The value is Attribute
publisher/NN has been applied with HE8, HE1.
at sentence 6, 5.
It has the total weight of 1.2
The value is Entity
publisher/NN location/NN has been applied with HE7, HA3.
at sentence 6, 6.
It has the total weight of -0.3
The value is Attribute
publisher/NN name/NN has been applied with HA8, HA3.
at sentence 6, 6.
It has the total weight of -1.7
The value is Attribute
quantity/NN of/IN books/NNS has been applied with HA2, HA7.
at sentence 4, 4.
It has the total weight of -1.2
The value is Attribute
specified/VBN in/IN has been applied with HR4.
at sentence 1.
It has the total weight of 0.8
The value is Relationship
stored/VBN for/IN has been applied with HR4.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
title/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
year/NN of/IN establishment/NN has been applied with HA3.
at sentence 6.
It has the total weight of -0.9
The value is Attribute
year/NN of/IN publication/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute

From the relationship record:

From the cardinality record:

## D.18. Machine

### The problem:

Machines manufacture components with identifying numbers. Each component has a name. Although a machine may manufacture many components, a single component may be produced on only one machine. A given component is stored in one or several warehouses. Each warehouse is managed by a warehouseman.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Machine | |
| Component | name |
| Warehouse | |
| warehouseman | |
| *Relationship*<br>Manufacture (machine, component)<br>Stored in (component, warehouse)<br>Managed by (warehouse, warehouseman) | |
| *Cardinality*<br>Many(machine, component)<br>One(component, machine)<br>Many(component, warehouse) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Machine | 11 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

### Overall result:

Recall: 100%
Precision: 92%

### Source of error:

a) Overgenerated relationship *produced on (component, machine)*

### Cases of multiple heuristics that are contradicting:

a) Component is correctly identified as an entity

component/NN has been applied with HE8, HA7, HE1.
at sentence 2, 4, 1.
It has the total weight of 0.7
The value is Entity

**The raw output from program:**

This is the output for file: machine.txt

The entity is component/NN
The attributes are
name/NN,

component/NN has been applied with
HE8, HA7, HE1.
at sentence 2, 4, 1.
It has the total weight of 0.7
The value is Entity
machine/NN has been applied with HE1,
HE1.
at sentence 3, 1.
It has the total weight of 1
The value is Entity
managed/VBN by/IN has been applied
with HR4.
at sentence 5.
It has the total weight of 0.8
The value is Relationship
manufacture/VB has been applied with
HR5.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
many/JJ has been applied with HC2.
at sentence 3.
It has the total weight of 0.9
The value is Cardinality
name/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
numbers/NNS has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
one/CD has been applied with HC4.
at sentence 3.
It has the total weight of 0.5
The value is Cardinality
one/CD or/CC has been applied with HC4.
at sentence 4.
It has the total weight of 0.5
The value is Cardinality
produced/VBN on/IN has been applied
with HR4, HR5.
at sentence 3, 3.
It has the total weight of 1.6
The value is Relationship
several/JJ has been applied with HC2.
at sentence 4.
It has the total weight of 0.9
The value is Cardinality
single/JJ has been applied with HC4.
at sentence 3.
It has the total weight of 0.5
The value is Cardinality
stored/VBN in/IN has been applied with
HR4, HR5.
at sentence 4, 4.
It has the total weight of 1.6
The value is Relationship
warehouse/NN has been applied with
HE1, HE1.
at sentence 5, 4.
It has the total weight of 1
The value is Entity
warehouseman/NN has been applied with
HE1.
at sentence 5.
It has the total weight of 0.5
The value is Entity
From the relationship record:
The relationship is stored/VBN in/IN
1st entity is component/NN
2nd entity is warehouses/NNS
at line 4
The relationship is produced/VBN on/IN
1st entity is component/NN
2nd entity is machine/NN
at line 3
The relationship is managed/VBN by/IN
1st entity is warehouse/NN
2nd entity is warehouseman/NN
at line 5
The relationship is manufacture/VB
1st entity is machine/NN
2nd entity is components/NNS
at line 3

From the cardinality record:
The cardinality is many
(component/NN,warehouses/NNS)
at line 4
The cardinality is many
(machine/NN,components/NNS)
at line 3
The cardinality is one
(component/NN,machine/NN)
at line 3

## D.19. Musician

### The problem:

Each musician that records has a name, an address and a phone number. Each instrument that is used in songs has a name and a musical key. Each album has a title, a copyright date, a speed and an album identifier. Each song has a title and an author. Each song is performed by one or more musicians, and a musician may perform a number of songs.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Musician | Name, address, phone number |
| Instrument | Name, musical key |
| Album | Title, copyright date, speed, album identifier |
| Song | Title, author |
| *Relationship*<br>Performed by(song, musician)<br>Used in (instrument, songs) | |
| *Cardinality*<br>Many (song, musician)<br>Many(musician, song) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Musician | 18 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

### Overall result:

Recall: 95%
Precision: 100%

### Source of error:

f) Undergenerated cardinality: many(musician,song)

### Cases of multiple heuristics that are contradicting:

g) Song title is correctly identified as an attribute

song/NN title/NN has been applied with HE7, HA3.
at sentence 4, 4.
It has the total weight of -0.3
The value is Attribute

## The raw output from program:

This is the output for file: musician.txt

The entity is song/NN
The attributes are
song/NN title/NN,author/NN,

The entity is album/NN
The attributes are
title/NN,copyright/NN
date/NN,speed/NN,album/NN identifier/NN,

The entity is musician/NN
The attributes are
name/NN,address/NN,phone/NN
number/NN,

The entity is instrument/NN
The attributes are
instrument/NN name/NN,key/NN,

address/NN has been applied with HA3.
at sentence 1.
It has the total weight of -0.9
The value is Attribute
album/NN has been applied with HE8.
at sentence 3.
It has the total weight of 0.7
The value is Entity
album/NN identifier/NN has been applied
with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
author/NN has been applied with HA3.
at sentence 4.
It has the total weight of -0.9
The value is Attribute
copyright/NN date/NN has been applied with
HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
instrument/NN has been applied with HE8.
at sentence 2.
It has the total weight of 0.7
The value is Entity
instrument/NN name/NN has been applied
with HA8, HA3.
at sentence 2, 2.
It has the total weight of -1.7
The value is Attribute
key/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
musician/NN has been applied with HE8, HE1.
at sentence 1, 5.
It has the total weight of 1.2
The value is Entity
name/NN has been applied with HA3.
at sentence 1.
It has the total weight of -0.9
The value is Attribute
number/NN has been applied with HEX.
at sentence 5.
It has the total weight of 100
The value is Non entity
one/CD or/CC has been applied with HC4.
at sentence 5.
It has the total weight of 0.5
The value is Cardinality
performed/VBN by/IN has been applied with
HR4, HR5.
at sentence 5, 5.
It has the total weight of 1.6
The value is Relationship
phone/NN number/NN has been applied with
HA8, HA3.
at sentence 1, 1.
It has the total weight of -1.7
The value is Attribute
song/NN has been applied with HE8, HE1.
at sentence 4, 2.
It has the total weight of 1.2
The value is Entity
song/NN title/NN has been applied with HE7,
HA3.
at sentence 4, 4.
It has the total weight of -0.3
The value is Attribute
speed/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9
The value is Attribute
title/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9
The value is Attribute
used/VBN in/IN has been applied with HR4.
at sentence 2.
It has the total weight of 0.8
The value is Relationship

From the relationship record:
The relationship is used/VBN in/IN
1st entity is instrument/NN
2nd entity is songs/NNS
at line 2
The relationship is performed/VBN by/IN
1st entity is song/NN
2nd entity is musicians/NNS
at line 5

From the cardinality record:
The cardinality is many
(song/NN,musicians/NNS)
at line 5

## D.20. Order

### The problem:

A person, identified by person id, can start an order. An order is identified by order id. Each order is started by one person, and one person can start more than one order. Each order is associated with one supplier. Each order is made up of a number of item, identified by item id. An item can appear in any number of orders. Item quantity is the quantity of items in a particular order. A quantity of any item in an order can be allocated to a project, identified by project number.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Person | Person id |
| Order | Order id |
| Item | Item id |
| Project | Project number |
| Supplier | |

*Relationship*
Start(person, order)
Made up(order, item)
Associated with(order, supplier)
Allocated to(order, project)

*Cardinality*
One(order, person)
Many(person, order)
Many (order, item)
Many (item, order)
One (order, supplier)

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Order | 18 | 0 | 2 | 2 | 0 | 0 | 0 | 1 |

### Overall result:

Recall: 100%
Precision: 82%

### Source of error:

a) Incorrect: *quantity* identified as entity
*b)* Incorrect: cardinality *one (person, order)*
c) Overgenerated: *item quantity*
d) Overgenerated: the relationship *started by(order, person)*

e) Wrongly attached: quantity in relationship *many (quantity, item)* –should be order

**Cases of multiple heuristics that are contradicting:**
Nil

**The raw output from program:**

This is the output for file: order.txt

The entity is order/NN
The attributes are
order/NN id/NN,

The entity is project/NN
The attributes are
project/NN number/NN,

The entity is item/NN
The attributes are
item/NN id/NN,

The entity is person/NN
The attributes are
person/NN id/NN,

allocated/VBN to/TO has been applied with HR4.
at sentence 8.
It has the total weight of 0.8
The value is Relationship
any/DT has been applied with HC2, HC2, HC2.
at sentence 6, 7, 8.
It has the total weight of 2.7
The value is Cardinality
associated/VBN with/IN has been applied with HR4, HR5.
at sentence 4, 4.
It has the total weight of 1.6
The value is Relationship
item/NN has been applied with HE9, HE1.
at sentence 5, 7.
It has the total weight of 1.2
The value is Entity
item/NN id/NN has been applied with HA8, HA2.
at sentence 5, 5.
It has the total weight of -1.5
The value is Attribute
item/NNP quantity/NN has been applied with HE7.
at sentence 7.
It has the total weight of 0.6
The value is Entity
made/VBN up/RP has been applied with HR4.
at sentence 5.
It has the total weight of 0.8
The value is Relationship
more/JJR than/IN has been applied with HC3.
at sentence 3.
It has the total weight of 0.6
The value is Cardinality
number/NN has been applied with HEX, HEX.
at sentence 5, 6.
It has the total weight of 200
The value is Non entity
one/CD has been applied with HC4, HC4.
at sentence 3, 4.
It has the total weight of 1
The value is Cardinality
order/NN has been applied with HE9, HE1.
at sentence 2, 6.
It has the total weight of 1.2
The value is Entity
order/NN id/NN has been applied with HA8, HA2.
at sentence 2, 2.
It has the total weight of -1.5
The value is Attribute
person/NN has been applied with HE9.
at sentence 1.
It has the total weight of 0.7
The value is Entity
person/NN id/NN has been applied with HA8, HA2.
at sentence 1, 1.
It has the total weight of -1.5
The value is Attribute
project/NN has been applied with HE9.
at sentence 8.
It has the total weight of 0.7
The value is Entity
project/NN number/NN has been applied with HA8, HA2.
at sentence 8, 8.
It has the total weight of -1.5
The value is Attribute
quantity/NN has been applied with HE1.

at sentence 7.
It has the total weight of 0.5
The value is Entity
start/VB has been applied with HR5.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
started/VBN by/IN has been applied with HR4, HR5.
at sentence 3, 3.
It has the total weight of 1.6
The value is Relationship
supplier/NN has been applied with HE1.
at sentence 4.
It has the total weight of 0.5
The value is Entity

From the relationship record:
The relationship is associated/VBN with/IN
1st entity is order/NN
2nd entity is supplier/NN
at line 4
The relationship is start/VB
1st entity is person/NN
2nd entity is order/NN
at line 3
The relationship is made/VBN up/RP
1st entity is order/NN
2nd entity is item/NN
at line 5
The relationship is started/VBN by/IN
1st entity is order/NN
2nd entity is person/NN
at line 3
The relationship is allocated/VBN to/TO
1st entity is order/NN
2nd entity is project/NN
at line 8

From the cardinality record:
The cardinality is many (quantity/NN,item/NN)
at line 8
The cardinality is many (item/NN,orders/NNS)
at line 6
The cardinality is one (person/NN,order/NN)
at line 3
The cardinality is one (order/NN,person/NN)
at line 3
The cardinality is many (quantity/NN,items/NNS)
at line 7
The cardinality is one (order/NN,supplier/NN)
at line 4

## D.21. Painter

### The problem:

```
A painter may paint many paintings. Each painting is painted by one
and only one painter. A painting may be exhibited in a gallery.
```

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Painter | |
| Painting | |
| Gallery | |
| *Relationship*<br><br>Paint (painter, paintings)<br>Exhibit (painting, gallery) | |
| *Cardinality*<br>One (painting, painter)<br>Many (painter, painting) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\_correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Painter | 6 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |

### Overall result:

Recall: 86%
Precision: 86%

### Source of error:

a) Overgenerated relationship: painted by (painter, paintings)
b) Undergenerated entity: exhibit(painting, gallery)

### Cases of multiple heuristics that are contradicting:

Nil

### The raw output from program:

This is the output for file: painter.txt

gallery/NN has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
many/JJ has been applied with HC2.
at sentence 1.
It has the total weight of 0.9
The value is Cardinality
one/CD has been applied with HC4.
at sentence 2.
It has the total weight of 0.5
The value is Cardinality
paint/VB has been applied with HR5.
at sentence 1.
It has the total weight of 0.8
The value is Relationship
painted/VBN by/IN has been applied with
HR4, HR5.
at sentence 2, 2.
It has the total weight of 1.6
The value is Relationship
painter/NN has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
painting/NN has been applied with HE1,
HE1.
at sentence 2, 1.
It has the total weight of 1
The value is Entity

From the relationship record:
The relationship is painted/VBN by/IN
1st entity is painting/NN
2nd entity is painter/NN
at line 2
The relationship is paint/VB
1st entity is painter/NN
2nd entity is paintings/NNS
at line 1

From the cardinality record:
The cardinality is many
(painter/NN,paintings/NNS)
at line 1
The cardinality is one
(painting/NN,painter/NN)
at line 2

## D.22. Photograph

### The problem:

Information about a collection of photographs is to be stored in a database. Photographs are identified by a photograph number and physical dimensions. They are provided on different types of papers, which has details of the weight of the paper and its finish. The photographers information to be recorded includes name, address and fax number. Sometimes photographs are out on loan to other departments.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Photograph | Photograph number, physical dimensions |
| Paper | Weight, finish |
| Photographer | Name, address, fax number |
| Department | |
| *Relationship*<br>Loan (photograph, department) | |
| *Cardinality* | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\_correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Photograph | 10 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |

### Overall result:

Recall: 83%
Precision: 83%

### Source of error:

a) Part_correct*: paper* identified as attribute
b) Incorrect: *loan* identified as entity
c) Undergenerated: relationship *loan(photograph, department)*

### Cases of multiple heuristics that are contradicting:

a) Details is correctly identified as a non-entity

details/NNS has been applied with HA3, HEX.
at sentence 3, 3.
It has the total weight of 99.1

The value is Non entity

b) Information is correctly identified as a non-entity

information/NN has been applied with HA7, HEX, HEX.
at sentence 1, 1, 4.
It has the total weight of 199.5
The value is Non entity

c) Paper is wrongly identified as an attribute

paper/NN has been applied with HA3, HE8.
at sentence 3, 3.
It has the total weight of -0.2
The value is Attribute

**The raw output from program:**

This is the output for file: photograph.txt

The entity is photographer/NN
The attributes are
name/NN,address/NN,fax/NN
number/NN,

The entity is types/NNS of/IN
papers/NNS
The attributes are
weight/NN,paper/NN,finish/NN,

The entity is photographs/NNP
The attributes are
photograph/NN
number/NN,dimensions/NNS,

address/NN has been applied with HA3.
at sentence 4.
It has the total weight of -0.9
The value is Attribute
database/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
departments/NNS has been applied with
HE1.
at sentence 6.
It has the total weight of 0.5
The value is Entity
details/NNS has been applied with HA3,
HEX.
at sentence 3, 3.
It has the total weight of 99.1
The value is Non entity
dimensions/NNS has been applied with
HA2.
at sentence 2.
It has the total weight of -0.7
The value is Attribute
fax/NN number/NN has been applied
with HA8, HA3.
at sentence 4, 4.
It has the total weight of -1.7
The value is Attribute
finish/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9
The value is Attribute
information/NN has been applied with
HA7, HEX, HEX.
at sentence 1, 1, 4.
It has the total weight of 199.5
The value is Non entity
loan/NN has been applied with HE1.
at sentence 6.
It has the total weight of 0.5
The value is Entity
name/NN has been applied with HA3.
at sentence 4.
It has the total weight of -0.9
The value is Attribute
paper/NN has been applied with HA3,
HE8.
at sentence 3, 3.
It has the total weight of -0.2
The value is Attribute
photograph/NN number/NN has been
applied with HA8, HA2.
at sentence 2, 2.
It has the total weight of -1.5
The value is Attribute
photographer/NN has been applied with
HE8.
at sentence 4.
It has the total weight of 0.7
The value is Entity

photographs/NNP has been applied with
HE9.
at sentence 2.
It has the total weight of 0.7
The value is Entity
weight/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9
The value is Attribute

From the relationship record:

From the cardinality record:

## D.23. Professor

### The problem:

Professors have an SSN, a name, an age, a rank and a research specialty. Projects have a project number, a sponsor name, a starting date, an ending date and a budget. Graduate students have an SSN, a student name, student age and a degree program. Each project is managed by one professor. Graduate students can work on multiple projects. A professor must supervise a graduate student on a project.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Professor | SSN, name,age,rank,research specialty |
| Project | Project number, sponsor name, starting date, ending date, budget |
| Graduate student | SSN, name,age,degree program |
| | |
| *Relationship*<br>Work on(graduate student, project)<br>Managed by(project, professor) | |
| *Cardinality*<br>One(project, professor)<br>Many(graduate student, project) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Professor | 21 | 0 | 0 | 0 | 1 | 0 | 2 | 0 |

### Overall result:

Recall: 96%
Precision: 100%

### Source of error:

*h)* Unattached: missing entities for relationship *work on(student, project)*
*i)* Undergenerated: missing entities for cardinality *many(student, project)*

### Cases of multiple heuristics that are contradicting:

a. Research specialty is correctly identified as an attribute

research/NN specialty/NN has been applied with HE7, HA3.
at sentence 1, 1.
It has the total weight of -0.3
The value is Attribute

b) Student age is correctly identified as an attribute

student/NN age/NN has been applied with HE7, HA3.
at sentence 3, 3.
It has the total weight of -0.3
The value is Attribute

**The raw output from program:**

This is the output for file: Professor.txt

The entity is project/NN
The attributes are
project/NN number/NN,

The entity is professors/NNS
The attributes are
sSN/NNP,name/NN,age/NN,rank/NN,re
search/NN specialty/NN,

The entity is projects/NNS
The attributes are
project/NN number/NN,sponsor/NN
name/NN,starting/NN
date/NN,ending/NN
date/NN,budget/NN,

The entity is graduate/NN students/NNS
The attributes are
sSN/NNP,student/NN
name/NN,student/NN
age/NN,degree/NN program/NN,

age/NN has been applied with HA3.
at sentence 1.
It has the total weight of -0.9
The value is Attribute
budget/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
degree/NN program/NN has been applied
with HE7, HA3.
at sentence 3, 3.
It has the total weight of -0.3
The value is Attribute
ending/NN date/NN has been applied
with HA8, HA3.
at sentence 2, 2.
It has the total weight of -1.7
The value is Attribute
graduate/NN student/NN has been
applied with HE7, HE7, HE8, HE7.
at sentence 6, 3, 3, 5.
It has the total weight of 2.5
The value is Entity
managed/VBN by/IN has been applied
with HR4, HR5.
at sentence 4, 4.
It has the total weight of 1.6
The value is Relationship
multiple/JJ has been applied with HC2.
at sentence 5.
It has the total weight of 0.9
The value is Cardinality
name/NN has been applied with HA3.
at sentence 1.
It has the total weight of -0.9
The value is Attribute
one/CD has been applied with HC4.
at sentence 4.
It has the total weight of 0.5
The value is Cardinality
professor/NN has been applied with HE1,
HE8.
at sentence 4, 1.
It has the total weight of 1.2
The value is Entity
project/NN has been applied with HE1,
HE8.
at sentence 4, 2.
It has the total weight of 1.2
The value is Entity
project/NN number/NN has been applied
with HA8, HA3.
at sentence 2, 2.
It has the total weight of -1.7
The value is Attribute
rank/NN has been applied with HA3.
at sentence 1.
It has the total weight of -0.9
The value is Attribute
research/NN specialty/NN has been
applied with HE7, HA3.
at sentence 1, 1.
It has the total weight of -0.3
The value is Attribute
sponsor/NN name/NN has been applied
with HA8, HA3.
at sentence 2, 2.
It has the total weight of -1.7
The value is Attribute
sSN/NNP has been applied with HA3,
HA3.
at sentence 1, 3.
It has the total weight of -1.8
The value is Attribute

starting/NN date/NN has been applied
with HA8, HA3.
at sentence 2, 2.
It has the total weight of -1.7
The value is Attribute
student/NN age/NN has been applied
with HE7, HA3.
at sentence 3, 3.
It has the total weight of -0.3
The value is Attribute
student/NN name/NN has been applied
with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
work/VB on/IN has been applied with
HR4, HR5.
at sentence 5, 5.
It has the total weight of 1.6
The value is Relationship

From the relationship recor:
The relationship is managed/VBN by/IN
1st entity is project/NN
2nd entity is professor/NN
at line 4

From the cardinality record:
The cardinality is one
(project/NN,professor/NN)
at line 4

## D.24. Project

### The problem:

Each employee belongs to a single department, and has one manager within that department. An employee can work on several projects. Each project has a start date, a finish date, and a team of employees assigned to it. One employee is assigned as the project manager. Projects are identified by a project code. Most projects are carried out for a single client, although there are internal projects for which there is no client. A client may undertake several projects. Clients are identified by a code.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Employee | |
| Department | manager |
| Project | Start date, finish date, team of employees, project manager, project code |
| Client | code |
| *Relationship*<br>Belongs to (employee, department)<br>Carried out (project, client)<br>Work on (employee, project) | |
| *Cardinality*<br>Many(employee, project)<br>One(project, client)<br>Many(client, project) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\_correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Project | 15 | 2 | 0 | 1 | 0 | 0 | 2 | 0 |

### Overall result:

Recall: 82%
Precision: 78%

### Source of error:

a) Part_correct: department identified as an attribute
b) Part_correct: project manager identified as an entity
c) Part_correct: employee identified as an attribute
d) Overgenerated: undertake(client, project)
e) Unattached: entities to belong to(employee, department)

### Cases of multiple heuristics that are contradicting:

a) Employee wrongly identified as an attribute

employee/NN has been applied with HE1, HA3, HA3.
at sentence 1, 3, 3.
It has the total weight of -1.3
The value is Attribute

**The raw output from program:**

This is the output for file: project.txt

The entity is project/NN
The attributes are
start/NN date/NN,finish/NN
date/NN,team/NN of/IN
employees/NNS,project/NN code/NN,

The entity is clients/NNS
The attributes are
code/NN,

belongs/VBZ to/TO has been applied with HR4, HR5.
at sentence 1, 1.
It has the total weight of 1.6
The value is Relationship
carried/VBN out/RP has been applied with HR4, HR5.
at sentence 6, 6.
It has the total weight of 1.6
The value is Relationship
client/NN has been applied with HE1, HE9.
at sentence 6, 8.
It has the total weight of 1.2
The value is Entity
code/NN has been applied with HA2.
at sentence 8.
It has the total weight of -0.7
The value is Attribute
department/NN has been applied with HA3.
at sentence 1.
It has the total weight of -0.9
The value is Attribute
employee/NN has been applied with HE1, HA3, HA3.
at sentence 1, 3, 3.
It has the total weight of -1.3
The value is Attribute
finish/NN date/NN has been applied with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
has/VBZ has been applied with HR5.
at sentence 1.
It has the total weight of 0.8
The value is Relationship
manager/NN has been applied with HA3.
at sentence 1.
It has the total weight of -0.9
The value is Attribute
one/CD has been applied with HC4.
at sentence 1.
It has the total weight of 0.5
The value is Cardinality
project/NN has been applied with HE8, HE1, HE9.
at sentence 3, 2, 5.
It has the total weight of 1.9
The value is Entity
project/NN code/NN has been applied with HA8, HA2.
at sentence 5, 5.
It has the total weight of -1.5
The value is Attribute
project/NN manager/NN has been applied with HE7.
at sentence 4.
It has the total weight of 0.6
The value is Entity
several/JJ has been applied with HC2, HC2.
at sentence 2, 7.
It has the total weight of 1.8
The value is Cardinality
single/JJ has been applied with HC4, HC4.
at sentence 1, 6.
It has the total weight of 1
The value is Cardinality
start/NN date/NN has been applied with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
team/NN of/IN employee/NN has been applied with HA3.
at sentence 3.
It has the total weight of -0.9
The value is Attribute
undertake/VB has been applied with HR5.
at sentence 7.
It has the total weight of 0.8
The value is Relationship
work/VB on/IN has been applied with HR4, HR5.
at sentence 2, 2.
It has the total weight of 1.6
The value is Relationship

From the relationship record:
The relationship is undertake/VB
1st entity is client/NN
2nd entity is projects/NNS
at line 7
The relationship is carried/VBN out/RP
1st entity is projects/NNS
2nd entity is client/NN
at line 6
The relationship is work/VB on/IN
1st entity is employee/NN
2nd entity is projects/NNS
at line 2

From the cardinality record:
The cardinality is many
(employee/NN,projects/NNS)
at line 2
The cardinality is many
(client/NN,projects/NNS)
at line 7
The cardinality is one
(projects/NNS,client/NN)
at line 6

## D.25. Reliable Rentals

### The problem:

Each outlet is allocated a stock of vehicles for hire. Individual vehicles may be moved between locations as required, but only the current location of each vehicle is stored. The registration number uniquely identifies each vehicle for hire and is used when hiring a vehicle to client.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Outlet | |
| Vehicles | Registration number |
| Location | |
| Hire | |
| Client | |
| *Relationship*<br>Moved (vehicles, location)<br>Allocated (outlet, vehicles) | |
| *Cardinality* | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Reliablerentals | 6 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |

### Overall result:

Recall: 75%
Precision: 75%

### Source of error:

a) Incorrect: *hiring* identified as an entity
b) Ask user: *vehicle*
c) Undergenerated: the relationship *allocated(outlet, vehicles)*
d) Unattached: *registration number* is not attached to *Vehicle*

### Cases of multiple heuristics that are contradicting:

b) Vehicle has an initial value of 'Ask user'

vehicle/NN has been applied with HE1, HA7.
at sentence 2, 2.
It has the total weight of 0
The value is Entity

**The raw output from program:**

This is the output for file: reliablerentals.txt

client/NN has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
hire/NN has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
location/NN has been applied with HE1,
HE1.
at sentence 2, 2.
It has the total weight of 1
The value is Entity
moved/VBN between/IN has been applied
with HR4.
at sentence 2.
It has the total weight of 0.8
The value is Relationship
outlet/NN has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
registration/NN number/NN has been
applied with HA8.
at sentence 3.
It has the total weight of -0.8
The value is Attribute
vehicle/NN has been applied with HE1,
HA7.
at sentence 2, 2.
It has the total weight of 0
The value is Entity

From the relationship record:
The relationship is moved/VBN
between/IN
1st entity is vehicles/NNS
2nd entity is locations/NNS
at line 2

From the cardinality record:

## D.26. Sales representatives

### The problem:

A database is to contain information concerning sales representatives, sales areas and products. Each sales representative is responsible for sales in one or more areas. Each area has one or more responsible sales representatives. Similarly, each sales representative is responsible for sales of one or more products, and each product has one or more sales representatives. Every product is sold in every area. However, no two sales representatives sell the same product in the same area. Every sales representative sells the same set of products in every area for which that sales representative is responsible.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Sales representative | |
| Product | |
| Area | |
| | |
| *Relationship*<br>has(sales representative, area)<br>has(sales representative, product) | |
| *Cardinality*<br>Many (sales representative, area)<br>Many (area, sales representative)<br>Many (sales representative, product)<br>Many (product, sales representative) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\_correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Salesrep | 9 | 0 | 1 | 0 | 0 | 0 | 0 | 6 |

### Overall result:

Recall: 100%
Precision: 90%

### Source of error:

a) Incorrect: sales identified as an entity
b) Wrongly attached: *sales* has been attached to all cardinalities instead of *sales representative*
c) Wrongly attached: sales has been attached to all relationships instead of *sales representative*

**Cases of multiple heuristics that are contradicting:**
Nil

**The raw output from program:**
This is the output for file: salesrep.txt

area/NN has been applied with HE8, HE1.
at sentence 3, 1.
It has the total weight of 1.2
The value is Entity
database/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
has/VBZ has been applied with HR5, HR5.
at sentence 3, 4.
It has the total weight of 1.6
The value is Relationship
information/NN has been applied with HEX.
at sentence 1.
It has the total weight of 100
The value is Non entity
one/CD or/CC has been applied with HC4, HC4, HC4.
at sentence 2, 3, 4.
It has the total weight of 1.5
The value is Cardinality
product/NN has been applied with HE8, HE1.
at sentence 4, 1.
It has the total weight of 1.2
The value is Entity
sales/NNS has been applied with HE1.
at sentence 2.
It has the total weight of 0.5
The value is Entity
sales/NNS representative/NN has been applied with HE7, HE7, HE7.
at sentence 2, 4, 7.
It has the total weight of 1.8
The value is Entity
sold/VBN in/RP has been applied with HR4.
at sentence 5.
It has the total weight of 0.8
The value is Relationship

From the relationship record:
The relationship is has/VBZ
1st entity is area/NN
2nd entity is sales/NNS
at line 3
The relationship is has/VBZ
1st entity is product/NN
2nd entity is sales/NNS
at line 4

From the cardinality record:
The cardinality is many
(product/NN,sales/NNS)
at line 4
The cardinality is many
(sales/NNS,areas/NNS)
at line 2
The cardinality is many
(sales/NNS,products/NNS)
at line 4
The cardinality is many
(area/NN,sales/NNS)
at line 3

## D.27. Student Hall

### The problem:

Students may rent a room in a university hall or student flat. Each hall has a name, address, telephone number and a hall manager. The halls provide only single rooms, which have a room number, place number and monthly rent rate.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Student | |
| Student flat | |
| Hall | Name, address, telephone number, hall manager |
| Room | Room number, place number, rent rate |
| *Relationship*<br>Rent (student, room)<br>Provide (hall, room) | |
| *Cardinality*<br>One (hall, room) | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Stud_hall | 13 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |

### Overall result:

Recall: 93%
Precision: 93%

### Source of error:

a) Wrongly attached: the entity *Hall* is wrongly attached to the attributes of *Room*
b) Overgenerated: the attribute *room number* attached to both *room* and *halls*
c) Undergenerated: the relationship rent(student, room)

### Cases of multiple heuristics that are contradicting:

a) Rent rate correctly identified as an attribute

rent/NN rate/NN has been applied with HE7, HA3.
at sentence 3, 3.
It has the total weight of -0.3
The value is Attribute

**The raw output from program:**

This is the output for file: student_hall.txt

The entity is hall/NN
The attributes are
name/NN,address/NN,telephone/NN
number/NN,hall/NN manager/NN,
The entity is room/NN
The attributes are
room/NN number/NN,

The entity is halls/NNS
The attributes are
room/NN number/NN,place/NN
number/NN,rent/NN rate/NN,

address/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
hall/NN has been applied with HE8, HE8.
at sentence 2, 3.
It has the total weight of 1.4
The value is Entity
hall/NN manager/NN has been applied
with HE7, HA3.
at sentence 2, 2.
It has the total weight of -0.3
The value is Attribute
name/NN has been applied with HA3.
at sentence 2.
It has the total weight of -0.9
The value is Attribute
place/NN number/NN has been applied
with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7
The value is Attribute
provide/VBP has been applied with HR5.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
rent/NN rate/NN has been applied with
HE7, HA3.
at sentence 3, 3.
It has the total weight of -0.3
The value is Attribute
room/NN has been applied with HE1,
HE1.
at sentence 1, 3.
It has the total weight of 1
The value is Entity
room/NN number/NN has been applied
with HA8, HA3.
at sentence 3, 3.
It has the total weight of -1.7

The value is Attribute
single/JJ has been applied with HC4.
at sentence 3.
It has the total weight of 0.5
The value is Cardinality
student/NN flat/NN has been applied
with HE7.
at sentence 1.
It has the total weight of 0.6
The value is Entity
students/NNS has been applied with HE1.
at sentence 1.
It has the total weight of 0.5
The value is Entity
telephone/NN number/NN has been
applied with HA8, HA3.
at sentence 2, 2.
It has the total weight of -1.7
The value is Attribute

From the relationship record:
The relationship is provide/VBP
1st entity is halls/NNS
2nd entity is rooms/NNS
at line 3

From the cardinality record:
The cardinality is one
(halls/NNS,rooms/NNS)
at line 3

## D.28. Student

### The problem:

Each student has a name, identification number, home address, term address and a number of qualifications for which the subject, grade and level are recorded. Each student is registered for one course where each course has a name and an identification number. Record is kept of the number of students registered for each course.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| student | Name, identification number, home address, term address |
| Course | Name, identification number |
| Qualification | Subject, grade, level |
| | |
| *Relationship*<br>Registered for(student, course) | |
| *Cardinality* | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\_correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Student | 12 | 1 | 0 | 0 | 0 | 0 | 0 | 4 |

### Overall result:

Recall: 92%
Precision: 92%

### Source of error:

a) Part_correct: *qualification* identified as an attribute
b) Wrongly attached: *qualifications, subject, grade* and *level* are wrongly attached to the entity *Student*

### Cases of multiple heuristics that are contradicting:

a) Student correctly identified as an entity

student/NN has been applied with HE8, HA7, HE1.
at sentence 1, 1, 3.
It has the total weight of 0.7
The value is Entity

b) Number correctly identified as non-entity

number/NN has been applied with HA3, HA7, HEX, HA7, HEX.
at sentence 1, 1, 1, 3, 3.
It has the total weight of 198.1
The value is Non entity

**The raw output from program:**

This is the output for file: student.txt

The entity is course/NN
The attributes are
name/NN,identification/NN number/NN,

The entity is student/NN
The attributes are
name/NN,identification/NN
number/NN,home/NN
address/NN,term/NN
address/NN,qualifications/NNS,subject/
NN,grade/NN,level/NN,

course/NN has been applied with HE8.
at sentence 2.
It has the total weight of 0.7
The value is Entity
grade/NN has been applied with HA3,
HA7.
at sentence 1, 1.
It has the total weight of -1.4
The value is Attribute
home/NN address/NN has been applied
with HA8, HA3, HA7.
at sentence 1, 1, 1.
It has the total weight of -2.2
The value is Attribute
identification/NN number/NN has been
applied with HA8, HA3, HA7, HA8, HA3.
at sentence 1, 1, 1, 2, 2.
It has the total weight of -3.9
The value is Attribute
level/NN has been applied with HA3,
HA7.
at sentence 1, 1.
It has the total weight of -1.4
The value is Attribute
name/NN has been applied with HA3,
HA7, HA3.
at sentence 1, 1, 2.
It has the total weight of -2.3
The value is Attribute

number/NN has been applied with HA3,
HA7, HEX, HA7, HEX.
at sentence 1, 1, 1, 3, 3.
It has the total weight of 198.1
The value is Non entity
qualifications/NNS has been applied with
HA3, HA7.
at sentence 1, 1.

It has the total weight of -1.4
The value is Attribute
registered/VBN for/IN has been applied
with HR4.
at sentence 3.
It has the total weight of 0.8
The value is Relationship
student/NN has been applied with HE8,
HA7, HE1.
at sentence 1, 1, 3.
It has the total weight of 0.7
The value is Entity
subject/NN has been applied with HA3,
HA7.
at sentence 1, 1.
It has the total weight of -1.4
The value is Attribute
term/NN address/NN has been applied
with HA8, HA3, HA7.
at sentence 1, 1, 1.
It has the total weight of -2.2
The value is Attribute

From the relationship record:
The relationship is registered/VBN for/IN
1st entity is students/NNS
2nd entity is course/NN
at line 3

From the cardinality record:

## D.29. Travel

### The problem:

A travel company arranges holiday for its customers. The company offers holidays in many countries. In each country, holidays are arranged in many different resorts and in each resort several different hotels are used. The company arranges charter flights to and from resorts. Each flight is assumed to go directly to the resort. The company has many agents and when a customer wants a holiday, he deals with an agent.

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Holiday | |
| Agent | |
| Customer | |
| Country | |
| Resort | |
| Flight | |
| Hotel | |

*Relationship*
Arranged (holiday, resort)
Arrange(flight, resort)
Deal(agent, customer)

*Cardinality*
Many(holidays, countries)
Many(resort,hotel)
Many(holiday, resort)

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\_correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Travel | 10 | 2 | 1 | 0 | 0 | 1 | 2 | 1 |

### Overall result:

Recall: 77%
Precision: 71%

### Source of error:

a) Part_correct: *agent* identified as attribute
b) Part_correct: *customer* identified as attribute
c) Ask user: *holiday* has an initial value of 'Ask user'
d) Incorrect: the business environment, *travel company*, identified as entity
e) Wrongly attached relationship , *assumed to,* to the entities *flight* and *resort*
f) Unattached entities to the relationship: deal(agent, customer)

**Cases of multiple heuristics that are contradicting:**

a) Holiday had an initial value of 'Ask user'

holiday/NN has been applied with HE1, HA3, HE1.
at sentence 1, 6, 2.
It has the total weight of 0.1
The value is Entity

b) Customer is wrongly identified as an attribute

customer/NN has been applied with HA3, HE1.
at sentence 6, 1.
It has the total weight of -0.4
The value is Attribute

**The raw output from program:**

This is the output for file: travel.txt

agent/NN has been applied with HA3, HA3.
at sentence 6, 6.
It has the total weight of -1.8
The value is Attribute
arranged/VBN in/IN has been applied with HR4, HR5.
at sentence 3, 3.
It has the total weight of 1.6
The value is Relationship
assumed/VBN to/TO has been applied with HR4.
at sentence 5.
It has the total weight of 0.8
The value is Relationship
charter/NN flight/NN has been applied with HE7.
at sentence 4.
It has the total weight of 0.6
The value is Entity
company/NN has been applied with HEX, HEX, HEX.
at sentence 2, 4, 6.
It has the total weight of 300
The value is Non entity
country/NN has been applied with HE1, HE1.
at sentence 3, 2.
It has the total weight of 1
The value is Entity
customer/NN has been applied with HA3, HE1.
at sentence 6, 1.
It has the total weight of -0.4
The value is Attribute
deals/VBZ with/IN has been applied with HR4.
at sentence 6.
It has the total weight of 0.8
The value is Relationship
flight/NN has been applied with HE1.
at sentence 5.
It has the total weight of 0.5
The value is Entity
has/VBZ has been applied with HR5.
at sentence 6.
It has the total weight of 0.8
The value is Relationship
holiday/NN has been applied with HE1, HA3, HE1.
at sentence 1, 6, 2.
It has the total weight of 0.1
The value is Entity
hotels/NNS has been applied with HE1.
at sentence 3.
It has the total weight of 0.5
The value is Entity
many/JJ has been applied with HC2, HC2, HC2.
at sentence 2, 3, 6.
It has the total weight of 2.7
The value is Cardinality
resort/NN has been applied with HE1, HE1.
at sentence 3, 3.
It has the total weight of 1
The value is Entity
several/JJ has been applied with HC2.
at sentence 3.
It has the total weight of 0.9
The value is Cardinality
travel/NN company/NN has been applied with HE7.

at sentence 1.
It has the total weight of 0.6
The value is Entity

From the relationship record:
The relationship is arranged/VBN in/IN
1st entity is holidays/NNS
2nd entity is resorts/NNS
at line 3
The relationship is assumed/VBN to/TO
1st entity is flight/NN
2nd entity is resort/NN
at line 5

From the cardinality record:
The cardinality is many
(holidays/NNS,countries/NNS)
at line 2
The cardinality is many
(resort/NN,hotels/NNS)
at line 3
The cardinality is many
(holidays/NNS,resorts/NNS)
at line 3

## D.30. University database

### The problem:

```
The university wishes to maintain a student database. The student will
be identified by a student registration number. In addition, the
student has name, name of school and name of study advisor to be
stored. Each module taught has its title, lecturer and room number.
```

### Actual solution:

| *Entity* | *Attribute* |
|---|---|
| Student | Student registration number, name, name of school, name of study advisor |
| Module | Title, lecturer, room number |
| *Relationship* | |
| *Cardinality* | |

### Result:

| *Dataset* | $N_{correct}$ | $N_{part\ correct}$ | $N_{incorrect}$ | $N_{overgenerated}$ | $N_{undergenerated}$ | $N_{ask}$ | $N_{unattach}$ | $N_{wrongattach}$ |
|---|---|---|---|---|---|---|---|---|
| Univ_d'base | 9 | 0 | 3 | 0 | 0 | 0 | 0 | 0 |

### Overall result:

Recall: 100%
Precision: 75%

### Source of error:

a) Incorrect: *student database* identified as entity
b) Incorrect: *university* identified as entity
c) Incorrect: relationship *wishes to(university, student)*

### Cases of multiple heuristics that are contradicting:

Nil

### The raw output from program:

This is the output for file:
university_database.txt

lecturer/NN has been applied with HA6.
at line 4.
It has the total weight of -0.9
The value is Attribute
module/NN has been applied with HE7.
at line 4.
It has the total weight of 0.7
The value is Entity
name/NN has been applied with HA6.
at line 3.
It has the total weight of -0.9
The value is Attribute
name/NN of/IN school/NN has been applied with HA6.
at line 3.
It has the total weight of -0.9
The value is Attribute

name/NN of/IN study/NN advisor/NN
has been applied with HA6.
at line 3.
It has the total weight of -0.9
The value is Attribute
room/NN number/NN has been applied
with HA11, HA6.
at line 4, 4.
It has the total weight of -1.7
The value is Attribute
student/NN has been applied with HE8,
HE7.
at line 2, 3.
It has the total weight of 1.4
The value is Entity
student/NN database/NN has been
applied with HE6.
at line 1.
It has the total weight of 0.6
The value is Entity
student/NN registration/NN number/NN
has been applied with HA11, HA5.
at line 2, 2.
It has the total weight of -1.5
The value is Attribute
title/NN has been applied with HA6.
at line 4.
It has the total weight of -0.9
The value is Attribute
university/NN has been applied with HE2.
at line 1.
It has the total weight of 0.5
The value is Entity
wishes/VBZ to/TO has been applied with
HR2.
at line 1.
It has the total weight of 0.8
The value is Relationship

The entity is module/NN
The attributes are
title/NN,lecturer/NN,room/NN
number/NN,

The entity is student/NN
The attributes are
student/NN registration/NN
number/NN,name/NN,name/NN of/IN
school/NN,name/NN of/IN study/NN
advisor/NN,

From the relationship record:
The relationship is wishes/VBZ to/TO
1st entity is university/NN
2nd entity is student/NN
at line 1

# Appendix E

# List of heuristics

## LIST OF HEURISTICS

| *Heuristic* | *Meaning* | *Weight* | *Status* |
|---|---|---|---|
| HEX | If a noun belongs to any of the set X where X= {record, database, company, system, information, organization, detail, interest, number, track} exclude it as a potential entity type candidate | 100 | Old |
| HE1 | All nouns are converted to entity types | 0.5 | Old |
| HE2 | A common noun may indicate an entity type | 0.5 | Old |
| HE3 | A proper noun may indicate an entity | 0.5 | Old |
| HE4 | A gerund may indicate an entity type which is converted from a relationship type | 0.5 | Old |
| HE5 | A clause may indicate a high-level entity type which hides a detailed Entity Relationship Diagram (ERD) (Chen, 1983) | 0.5 | Old |
| HE6 | If a noun occurs before a genitive, it may indicate an entity type | 0.5 | Old |
| HE7 | If compound nouns are present, check the last noun. If it is not one of the words in set S where S={number, no, code, date, type, volume, birth, id, address, name}, most likely it is an entity type. Else it may indicate an attribute type. | 0.5 | New |
| HE8 | If a noun occurs before the verb 'has'/ 'have', it may indicate an entity type | 0.7 | New |
| HE9 | If a noun occurs before the verb 'identified by', it may indicate an entity type | 0.7 | New |
| HA1 | A noun which takes the general form of TERM_SUFFIX such as noun_id, noun_no, noun_type or noun_number may indicate an attribute type | -0.9 | Old |
| HA2 | A noun phrase which follows the phrase "identified by" may indicate the presence of attribute types | -0.7 | Old |
| HA3 | A noun phrase succeeding the "has/have" verb phrase may indicate the presence of attribute types | -0.9 | Old |

| *Heuristic* | *Meaning* | *Weight* | *Status* |
|---|---|---|---|
| HA4 | An intransitive verb may indicate an attribute type | -0.4 | Old |
| HA5 | An adjective can be an attribute type | -0.3 | Old |
| HA6 | Genitive case in the noun phrase may indicate an attributive function | -0.3 | Old |
| HA7 | A noun phrase which precedes the verb phrase "is/are stored", "is/are recorded" or "is/are kept" or the phrase "is/are of interest" may indicate the presence of attribute types | -0.5 | New |
| HA8 | If a noun is followed directly by another noun and the latter belongs to set S where S={number, no, code, date, type, volume, birth, id, address, name}, this may indicate that both words are an attribute. Otherwise, it is most likely an entity (HE7). | -0.8 | New |
| HC1 | A noun or a prepositional phrase whose noun is singular gets a minimal and maximum cardinality of 1 | 0.5 | Old |
| HC2 | The adjective "many" or "any" may suggest a maximum cardinality | 0.9 | New |
| HC3 | A comparative adjective "more" followed by the preposition "than" and a cardinal number may indicate the degree of the cardinality between two entities | 0.6 | New |
| HC4 | Cardinal number "one" or the adjective "single" may indicate cardinality of one | 0.5 | New |
| HC5 | The noun "none" or the cardinal number "zero" may indicate the lower bound of a cardinality | 0.5 | New |
| HC6 | The phrase "one or more" or the adjective "multiple" may indicate a maximum cardinality. | 0.5 | New |
| HR1 | An adverb can indicate an attribute for relationship | 0.5 | Old |
| HR2 | A transitive verb can be a candidate for relationship type (Chen, 1983) | 0.5 | Old |
| HR3 | The preposition "for" can indicate a relationship type | 0.3 | New |

| *Heuristic* | *Meaning* | *Weight* | *Status* |
|---|---|---|---|
| HR4 | *A verb followed by a preposition such as "on", "in", "by" and "to" may indicate a relationship type* | *0.8* | *New* |
| HR5 | A verb that appears before an adjective "many" or "any" may indicate a relationship | 0.8 | New |

# Appendix F

# References

**REFERENCES**

Allen, J. (1995) Natural Language Understanding. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California.

Anderson, J.R. (1988) *The Expert Module*. In M.C. Polson and J.J Richardson (Eds.), Foundations of Intelligent Tutoring Systems. pp. 21-53. New Jersey: Lawrence Erlbaum Associates, Inc.

Antony, S.R. and Batra, D. (2002) *Codasys: A Consulting Tool for Novice Database Designers*, ACM SIGMIS Database, Vol. 33, Issue 3, pp. 54-68.

Batra, D. and Antony, S.R. (1994) *Novice Errors in Conceptual Database Design*. European Journal of Information Systems. Vol.3, 1, pp. 57-94.

Batra, D., Hoffer, J.A. and Bostrom, R.P. (1990) *Comparing Representations with Relational and EER Models*. Communications of the ACM, Vol. 33, pp. 126-139.

Batra, D. and Zanakis, S.H. (1994) *A Conceptual Database Design Methodology Based on Rules and Heuristics*, European Journal of Information Systems, Vol. 3, No. 3, pp. 228-239.

Boehm, B.W. (1981) Software Engineering Economics, Prentice-Hall, Inc., Englewood Cliffs, NJ, USA.

Booch, G., Rumbaugh, J. and Jacobson, I. (1999) The Unified Modeling Language User Guide, Addison-Wesley.

Brill, E. (1992) *A Simple Rule-Based Part of Speech Tagger*. In Proceedings of the Third Conference on Applied Natural Language Processing, ACL, Italy, pp.152-155.

Brill, E. (1994) *Some Advances in Transformation-based Part-of-speech Tagging,* In Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94), August 1-4, 1994, Seattle, Washington.

Brown, J.S., Burton, R.R., and deKleer, J. (1982) *Pedagogical, Natural Language and Knowledge Engineering Techniques in SOPHIE I, II and III*. In D. Sleeman and J.S Brown (Eds.), Intelligent Tutoring Systems (pp. 227-282). New York: Academic Press.

Buchanan and Shortliffe (eds.) (1984): Rule-based Expert Systems: The MYCIN Experiments of the Standford Heuristic Programming Project, Addison-Wesley.

Buchholz, E., Cyriaks, H., Dusterhoft, A., Mehlan, H., and B. Thalheim. (1995) *Applying a Natural Language Dialogue Tool for Designing Databases*. In Proceedings of the First Workshop on Applications of Natural Language to Databases (NLDB'95), Versailles, France, pp. 119- 133.

Burg, J.F.M. and van de Riet R.P., (1995) *COLOR-X: Linguistically-based Event Modeling: A General Approach to Dynamic Modeling*, in Proc. of Advanced Information Systems Engineering (CAISE'95), J. Iivari, K. Lyytinen and M. Rossi (eds.) (LNCS-932), Springer-Verlag, Berlin, pp. 26-39.

Burg, J.F.M. and Van de Riet R.P. (1996) *A Natural Language and Scenario based Approach to Requirements Engineering*. In Natuerlichsprachlicher Entwurf von Informationssystemen (NEI'96), Tutzing, Germany.

Burg, J.F.M and van de Riet, R.P. (1998) *Using Knowledge from WordNet for Conceptual Modelling*. In Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database, Cambridge, MA: MIT Press.

Burns, H. L. and Capps, C. G. (1988) Foundations of Intelligent Tutoring Systems: An Introduction, Foundations of Intelligent Tutoring Systems, Lawrence Erlbaum Associates, Hillsdale, NJ, USA.

Burton, R.R. (1982). *Diagnosing Bugs in a Simple Procedural Skill*. In D. Sleeman and J.S Brown (Eds.), Intelligent Tutoring Systems (pp.157-183). New York: Academic Press.

Burton, R.R., and Brown, J.S. (1982) *An investigation of computer coaching for informal learning activities*. In D. Sleeman and J.S Brown (Eds.), Intelligent Tutoring Systems (pp. 79-98). New York: Academic Press.

Canavan, B. (1996) *An Intelligent Tutoring System for Database Design*. Masters Dissertation, Faculty of Informatics, University of Ulster.

Carpenter, D.A. (1992) *Are We Teaching Database Design Properly?*, Journal of Computer Information Systems, pp. 9-12.

Cerpa, N. (1995) *Pre-physical Database Design Heuristics*, Information and Management, Vol. 28, pp. 351-359.

Chen, P. (1976) *The Entity-relationship Model: Toward A Unified View of Data*, ACM Transactions on Database Systems, 1(1), pp. 9-36.

Chen, P.P. (1983) *English Sentence Structure and Entity-Relationship Diagram,* Information Sciences, Vol. 1, No. 1, Elsevier, pp. 127-149.

Chen, P.P. (1998) *From Ancient Egyptian Language to Future Conceptual Modeling in: Conceptual Modeling: Current Issues and Future Directions*, Chen, P.P., (eds.), Springer-Verlag, Berlin, Lecturing Notes in Computer Sciences, No. 1565, pp. 57-66.

Clancey, W.J. (1982) *Tutoring Rules for Guiding a Case Method Dialogue*. In D. Sleeman and J.S Brown (Eds.), Intelligent Tutoring Systems (pp. 201-225). New York: Academic Press.

Connolly, T. and Begg, C. (1999) Database Systems: A Practical Approach to Design, Implementation, and Management, 2nd Edition, Addison-Wesley.

Constantino-Gonzalez, M. and Suthers, D.D. (2000) *A Coached Collaborative Learning Environment for Entity-Relationship Modeling*. Proceedings of the 5th International Conference (ITS 2000), G. Gauthier, C. Frasson, and K. VanLehn (Eds.), pp. 325-333, Berlin: Springer-Verlag.

Daelemans, W., Zavrel, J., Berck, P. and Gillis, S. (1996) *MBT: A Memory-based Part of Speech Tagger Generator*, In Proceedings of the Fourth Workshop on Very Large Corpora, Copenhagen, pp. 14-27.

Daelemans, W. and van de Bosch, A. (1992) *Generalisation Performance of Backpropagation Learning on a Syllabification Task.* In M.F.J. Drossaers and A. Nijholt (Eds.), Proceedings of TWLT3: Connectionism and Natural Language Processing, pp. 27-37, Enschede, Twente University.

Daelemans W., van den Bosch, A., Zavrel, J., Veenstra, J., Buchholz, S., and Busser, B. (1998) *Rapid Development of NLP Modules With Memory-based Learning*. In Proceedings of ELSNET in Wonderland, Utrecht, pp. 105-113.

Daelemans, W., Zavrel, J., van der Sloot, K. and van de Bosch, A. (2000) *TiMBL: Tilburg Memory Based Learner,* Version 3.0, Reference Guide, ILK Technical Report 00-01.

Dean, C. and Whitlock, Q. (1992) *A handbook of computer-based training*, 3rd ed., Gulf Publishing Company.

Dullea, J., Yeolsong, I. And Lamprou, I. (2003) *An Analysis of Structural Validity in ER Modeling*, Data and Knowledge Engineering, Vol. 47, Issue 2, pp. 167-205.

Eick, C. F.and Lockemann, P.C. (1985) *Acquisition of Terminology Knowledge Using Database Design Techniques*. In Proceedings ACM SIGMOD Conference, pp. 84-94, Austin, USA.

Elmasri, R. and Navathe, S.B. (2004) Fundamentals of Database Systems, 4th Edition, Addison-Wesley, London.

Fellbaum, C.A. (1998) *WordNet: An Electronic Lexical Database*, Fellbaum, C. (Ed.). Cambridge, MA: MIT Press.

Fillmore, C.J. (1971) Some Problems for Case Grammar. In R.J. O'Brien (ed), 22nd Annual Round Table. Linguistics: Developments of the Sixties-viewpoints of the Seventies, Vol. 24 of Monograph Series on Language and Linguistics. Georgetown University Press, Washington D.C., pp. 35-36.

Fuchs, N.E. and Schwitter, R. (1996), *Attempt to Controlled English* (ACE), CLAW 96, First International Workshop on Controlled Language Applications, University of Leuven, Belgium.

Fuchs, N.E. and Schwertel, U. (2003) *Reasoning in Attempto Controlled English*, in: F. Bry, N. Henze and J. Maluszynski (eds.): *Principles and Practice of Semantic Web Reasoning*, International Workshop, PPSWR 2003, Mumbai, India, December 2003. Lecture Notes in Computer Science 2901, Springer Verlag, 2003.

Gildea, D. and Jurafsky, D. (2002) *Automatic Labeling of Semantic Roles*, Computational Linguistics, 28(3), pp. 245-288.

Griffin, D. and Tversky, A. (2002) *The Weighing of Evidence and the Determinants of Confidence*. In Gilovich, H., Griffin, D. and Kahneman, D. (Eds.), Heuristics and Biases: The Psychology of Intuitive Judgment, pp. 230-249, Cambridge University Press.

Grishman, R. and Sundheim, B., 1996 *Message Understanding Conference-6: A Brief History.* In Proceedings of the 16th International Conference on Computational Lingusitics, Copenhagen, pp. 466-471.

Gomez, F., Segami, C. and Delaune, C. (1999) *A System for the Semi-Automatic Generation of E-R Models from Natural Language Specifications,* Knowledge and Data Engineering, Vol. 29, pp.57-81.

Groner, M., Groner, R. and Bischof, W.F. (1983) *Approaches to Heuristics: A Historical Review.* In Groner, M., Groner, R. and Bischof, W.F.(Eds.), Methods of Heuristics, pp. 1-20, Lawrence Erlbaum Associates.

Halff, H.M. (1988) *Curriculum and Instruction in Automated Tutors*. In M.C. Polson and J.J Richardson (Eds.), Foundations of Intelligent Tutoring Systems, pp. 79-108, New Jersey: Lawrence Erlbaum Associates, Inc.

Harmain, H.M. (2000) *Building Object-oriented Conceptual Models Using Natural Language Processing Techniques,* Ph.D. Thesis, Department of Computer Science, University of Sheffield.

Harmain, H.M. and Gaizauskas, R. (2003) *CM-Builder: A Natural Language-Based CASE Tool for Object-Oriented Analysis*, Automated Software Engineering, Vol. 10, pp. 157-181.

Hirschman, L. and Thompson, H.S. (1995) Chapter 13. *Evaluation: Overview of Evaluation in Speech and Natural Language Processing*. In R.A. Cole, J. Mariani, H. Uszkoreit, A. Zaenen and V. Zue, (eds), Survey of the State of the Art in Human

Language Technology. Available at: http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html.

Jackendoff, R. (1983) Semantics and Cognition, The MIT Press, Cambridge, MA.

Jones, C.B. (1990) Systematic Software Development Using VDM, Prentice Hall International.

Jurafsky, D. and Martin, J. H. (2000) Speech and Language Processing, Prentice-Hall, New Jersey.

Kahneman, D., Slovic, P. and Tversky, A. (eds.) (1982): Judgement Under Uncertainty: Heuristics and Biases, Cambridge, University Press.

Kleen, B.A. (1993) *Are We Missing the Boat When Teaching Database Concepts and Applications?,* Journal of Computer Information Systems, pp.1-6.

Lester, J.C, Converse, S.A., Stone, B.A., Kahler, S.E., and Bhogal, R.S. (1997a) *The Persona Effect: Affective Impact of Animated Pedagogical Agents,* Proceedings of CHI'97, pp. 359-366.

Marsden, P. and Staniforth, L. (1996) *MERMott - A Multimedia Based Tool Supporting the Teaching of Entity-Relationship Modelling Within a Framework of Structured Systems Analysis*. Association for Information Systems, Americas Conference, Phoenix, Arizona, USA.

Martinez, P. and Garcia-Serrano, A. (2001) *On the Automatization of Database Conceptual Modelling through Linguistic Engineering.* Lecture Notes in Computer Science, 1959, pp. 276-287.

Meziane, F. (1994) From English to Formal Specifications, PhD Thesis, Department of Mathematics and Computer Science, University of Salford, UK.

Meziane, F. and Vadera, S. (2004) *Obtaining E-R Diagrams Semi-Automatically From Natural Language Specifications*, In proceedings of the 6th International Conference of Entreprise Information Systems, ICEIS 2004, 14 April –17April 2004, Portugal, pp. 638-642.

Miller, J.R. (1988) *The Role of Human-Computer Interaction in Intelligent Tutoring Systems*. In M.C. Polson and J.J Richardson (Eds.), Foundations of Intelligent Tutoring Systems, pp. 143-189. New Jersey: Lawrence Erlbaum Associates, Inc.

Mitrovic, A. (1998). *A Knowledge-Based Teaching System for SQL*, Proc. ED-MEDIA/ED-TELECOM'98, Freiburg, pp. 1027-1032.

Mitrovic, A. and Ohlsson, S. (1999) *Evaluation of A Constraint-Based Tutor for A Database Language*. International Journal of AI in Education, No.10, pp. 238-256.

Moody, D. (1996) *Graphical Entity Relationship Models: Towards a More User Understandable Representation of Data*. In B.Thalheim (ed), 15th International Conference on Conceptual Modelling, Cottbus, Germany. Lecture Notes in Computer Science, No.1157, pp. 227-244.

Munoz, M., Punyakanok, V., Roth, D. and Zimak, D. (1999) *A Learning Approach to Shallow Parsing*. EMNLP-VLC, the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, pp. 168-178.

Neri, F. and Saitta, L. (1997) *Machine Learning for Information Extraction.* In: Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, Maria Teresa Pazienza (Ed.), Springer-Verlag, Berlin, Lecture Notes in Artificial Intelligence, No. 1299, pp. 171-191.

Ohlsson, S. (1992) *Constraint-based Student Modeling*, Journal of Artificial Intelligence in Education, 2(3), pp. 5-18.

Raguphati, V., and Schkade, L.L. (1992) *An Intelligent Tutoring System for Database Design*. Proceedings of the Decision Sciences Institute Meeting, Vol.2, pp.603-605, San Francisco, USA.

Ramakrishnan, R. (1998) *Database Management Systems,* WCB/McGraw-Hill.

Ratnaparkhi, A. (1996) *A Maximum Entropy Part-of-speech-tagger,* In Proceedings of the Conference on Empirical Methods in Natural Language Processing, May 17-18, 1996, University of Pennsylvania.

Rolland, C., Proix, C. (1992) *A Natural Language Approach for Requirements Engineering*. In P. Loucopoulos, (editor), Proceedings of the Fourth International Conference CAiSE'92 on Advanced Information Systems Engineering, Volume 593 of Lecture Notes in Computer Science, Manchester, United Kingdom, SpringerVerlag, pp. 257—277.

Rosenthal, A. and Reiner, D. (1994) *Tools and Transformations-Rigorous and Otherwise-for Practical Database Design*, ACM Transactions on Database Systems, Vol. 19, No. 2, pp. 167-211.

Self, J., (ed.), (1988) Artificial Intelligence and Human Learning: Intelligent Computer Aided Instruction, Chapman and Hall, London.

Siemer, J., and Angelides, M.C. (1998). *A Comprehensive Method for the Evaluation of Complete Tutoring Systems*, Decision Support Systems, Vol. 22, pp. 85-102.

Sleeman, D.H, and Brown, J., (eds.) (1982) *Intelligent Tutoring Systems*. Academic Press.

Sleeman, D.H. (1982) *Assessing Competence in Basic Algebra*. In D. Sleeman and J.S Brown (Eds.), Intelligent Tutoring Systems (pp. 185-199). New York: Academic Press.

Somers, H. (2000) *Machine Translation*. In R. Dale et al. (eds), Handbook of Natural Language Processing, Marcel Dekker, Inc., New York, pp. 329-346.

Stankov, S. (1996) *Student Modeling Developing for Intelligent Tutoring Systems*. International Journal for Engineering Modelling, Vol. 9, pp. 35-41.

Steetskamp, R. (1995) *An Implementation of a Probabilistic Tagger*, Master's Thesis, TOSCA Research Group, University of Nijmegen, Nijmegen, The Netherlands.

Storey, V.C. (1988) *View Creation: An Expert System for Database Design*, Ph.D. Thesis, Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, Canada. Washington, D.C.: ICIT Press.

Storey, V.C. (1993) *A Selective Survey of the use of Artificial Intelligence for Database Design Systems*, Data and Knowledge Engineering, Vol. 11, pp. 61-102.

Storey, V.C. and Goldstein, R.C. (1988) *A Methodology for Creating user Views in Database Design*. ACM Transactions on Database Systems, Vol.13, pp.305-338.

Sugumaran, V. and Storey, V. (2002) *Ontologies for Conceptual Modeling: Their Creation, Use and Management*, Data and Knowledge Engineering, Vol. 42, pp. 251-271.

Suraweera, P. and Mitrovic, A. (2002) *KERMIT: A Constraint-based Tutor for Database Modeling.* In: S. Cerri, G. Gouarderes and F. Paraguacu (eds.), Proc. 6th Int. Conf on Intelligent Tutoring Systems, ITS 2002, Biarritz, France, LCNS 2363, pp. 377-387.

Tapanainen, P. (1996). *The Constraint Grammar Parser CG-2*. Publication of the University of Helsinki, No. 27, Helsinki, Finland.

Tjoa, A.M and Berger, L. (1993) *Transformations of Requirements Specifications Expressed in Natural Language into an EER Model*. Proceeding of the 12th International Conference on Approach, Airlington, Texas, USA, pp. 206-217.

Tong, A.K.Y. (1997) *Developing A Model for Tutoring Strategy Selection in Intelligent Tutoring Systems,* Ph.D Thesis, London School of Economics and Political Science.

Tseng, F.S.C, Chen, A.L.P. and Yang, W. (1992) *On Mapping Natural Language Constructs into Relational Algebra Through E-R Representation,* Data and Knowledge Engineering, No. 9, pp. 97-118.

VanLehn, K. (1988) *Student Modelling*. In M.C. Polson and J.J Richardson (Eds.), Foundations of Intelligent Tutoring Systems, pp. 55-78. New Jersey: Lawrence Erlbaum Associates, Inc.

Willits, J. (1992) Database Design and Construction: An Open Learning Course for Students and Information Managers, Library Association Publishing Ltd., London.

Zanakis, S.H. and Evans, J.R. (1981) *Heuristic 'Optimization': Why, When and How to use it.* Interfaces 11(5), pp. 84-91.

Zavrel, J. and Daelemans, W. (1999) *Recent Advances in Memory-based Part-of-Speech Tagging*, In Actas del VI Simposio Internacional de Comunicacion Social, Santiago de Cuba, pp. 590-597.

Zavrel, J. and Daelemans, W. (2003) *Feature-rich Memory-based Classification for Shallow NLP and Information Extraction,* In J. Franke, G. Nakhaeizadeh and I. Renz (Eds.), Text Mining, Theoretical Aspects and Applications, Springer Physica-Verlag, pp. 33-54.