Interim Report

By: Xiaofeng Fu

Supervisor: Anthony Hunter

Used project title: London Trip Information Advisor Current project title: London Restaurant Advisor

During the implementation of the project, I have changed the project title since the previous one, London Trip Information Advisor, involves many fields and it is difficult for me to implement all of them before the deadline of the project. Therefore, now I only focus on providing advises of London restaurants in my project.

Progress:

In the November project plan, I set several requirements in the MoSCoW table:

- A valid source of London trip information
- The ability to identify and understand users' questions
- The ability to find the potential answer from the source.
- The ability to translate the potential answer to an answer written in natural language.
- A website user interface to contain this system.
- Put the website on a server.
- A beautiful website user interface.

Until now, most of the requirements are implemented:

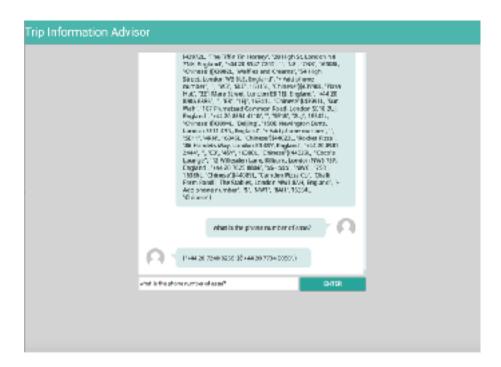
 To get sufficient restaurants data, I write a web scrawler to collect more than 18,000 sets of London restaurants information in a mySQL database from the website of TripAdvisor.

100	Name of the last o	V100	4000	4400.00		W 100	
A CHARLESON	CONTRACTOR CONTRACTOR PROPERTY.			-	-	-	100
T-LESCHOOL STATE OF THE PARTY O	THE REAL PROPERTY AND PERSONS ASSESSED.	ALC: UNKNOWN	10.00	900	100		THE SALE
T / T POR MOO	TERRODIC TO LANGE THE STATE COMME	ALC: UNKNOWN	20.00	7073	-	-	Section .
# 1 T SPECIMEN	TORONOUS COLUMNS (MEDICAL PROPER)	40.0449	0.00	900	-		Market restu
***********	CONTRACTOR OF THE PARTY AND ADDRESS.			-	-	-	-
manufacture 1	CONTRACTOR OF CONTRACTOR PROPERTY.				-	-	Page 100 (100 to 100 to
WE WANTED TO STATE OF THE PARTY.	SECURITY STATE OF THE PARTY AND THE		-	man .	100,000	-	
To have been by \$100m.	STREET, SHARL SHARL STREET, SHARL	ALC: NO.	10.00	900	1540		888
The Disserver Seather See \$ 600 has		ALC: UNKNOWN	10.00	WHITE CO.	1540		English Co.
The Street on Section See \$ 6500 at	Street on Street Code Street Street	AND DESCRIPTION	0.00	958	156		SAME THAT TO SEE
	THE REST, SHARPER WAS LOND ON THE PARTY			-	-	-	-
W1 (0.00 (0.00 (0.00)	THE ROOM, LINES CO., MICH., LANSING SHIP STREET,			-	-		***
TO THE MANAGE	OF THE ARRANGE THE LAST SECURIOR	ALC: UNKNOWN	20.00	100	100		No. Contraction
To the mount	HET THE WINNING THE LAST SET WITHOUT	ALC: UNKNOWN	M-90		100		Registration Country
To Discharge St.	HER THE HER HER THE LAST SEC WITHOUT	443 07680	10.00	Sec.	140		Steph Sterne
	THE REST, MARRIED WAY, LINE WAS THE PARTY			-	1991		Page 100 (100 to 100 to
Contraction of the Contraction o	NOTES AND THE WHILE WHEN PARTY AND ADDRESS.	-		-	1918	100	-
SURE PRODUCTS	Married Street, Page Stright, John St. Hollycon.			-	1991		1.0000
Long Section 1	METALS BOAT AND WHILE AND A PERSONNEL	ALC: NO RESIDENCE	20.00		100	4	NAME .
1,000 Research 1	METWOON STORY AND STORY AND ASSESSED ASSESSED.	ALC: NO PERSONS	M-10		100		Million Section 1
Long Springer	METERS BOAT AND MIGHT AND AT PERSON	ALC: NO SECURE	H- 60		100	4	Registration frameton
Commence of the Commence of th	Market State of the State of th			-	17.0	-	The same
Springer Ministration of the Control	Married State of the State of t			-	1999	-	100
Long Section 1	METALS STATE AND STATE AND ADDRESS OF PERSONS	ALC: WHEN	10.00		100		State State Street,
Sky St Therborne	ET Broker Strate Std CT Expert	ALC: UK THE	m - m		179		Married World
St STREET, ST	ET BANKSTON ON THE PROPER	40.00	M-00	940	179		Self-peak-
	the property of the page 1			-	110	100	NAME OF TAXABLE PARTY.
	The second section of the second section of the second section			-	-		
Mr. Street, St. Co.	the British was not the party of			-		-	The second
St. ST. State and St.	TO SHARE THE ARE THE PERSON.	ALC: UNKNOWN	-	-	778		Build Say Street
Negra	TO REPORT AND THEIR READING VISION SHIP THE EXPLANA	AND THE RES	m m	SM1.	1769		Married World
Name of the last o	TO REPORT THAT THEY READING WHEN BEY THE EXCENT	ALC: THE REAL PROPERTY.	0.00	987	756		449
mage.	commence and committee on the column		-	-	1988	-	
10001	CONTRACTOR AND ADDRESS OF THE PARTY.			9000	1990	-	100
Telluria.	STREET, SHE THE BROWN HARDEN TO SERVE	ALC: YES	10.00	BMT .	156		State Street
Personal Indiana	Tribution St. Look SCHOOL Expent	417 140 88	-	80.0	1798		Service.
PERSONAL PROPERTY.	THE REAL PROPERTY AND ADDRESS OF THE PERSON NAMED IN COLUMN TWO PERSONS AND ADDRESS OF THE PERSON NAMED IN COLUMN TWO PERSONS AND ADDRESS OF THE PERSON NAMED IN COLUMN TWO PERSON NAMED IN COLUMN TRANSPORT NAMED IN COLUMN TWO PERSON NAMED	44314088	-	814	1756		8100
	Comment of Comment of Street, Supplement		-	10.75	-	-	
	Commence of Commence of Street, St. Street,		-	10.7		-	-

- I write a parser to analyse users' questions and find the potential answer from the database. This parser includes four parts. The first part is query formulation. For this step, the system will extract all potential keywords from the query. Firstly, the system will ignore all stopwords and punctuations in the query. Secondly, entities will be tagged based on the database. Then the system will use POS-tagger in NLTK to tag every word in noun phrases to separate all adjective and adverb from the noun. The second part is answer type classification. Because I am focusing on a closed domain QA system, question types are fixed and it is not difficult to estimate what questions will be asked by users. The parser uses Jaccard Coefficient mechanism to implement this part. This mechanism is using a equation: probability = (Interaction of QueryTerms and user question) / (Union of QueryTerms and user questions) Firstly, I wrote a survey to collect potential questions from my friends. Then, I built a database to save the type of question and the related answer type in a special form. And then, I built a function to compare each word of user question with each query terms to find the number of word interaction. The query term with the highest probability is the type of the user question. For example, 'what is the best Burger King in London', the query term 'what is the adjective entity in loc' will be matched and the answer type will be 'description or definition'. Then it will be easier for me to determine the SQL query. The third part is relation extraction. The parser bind each keyword with its description in a tuple, ('Burger King', 'Entity'), using a external dictionary and the restaurant database. The forth part is SQL query generation. Based on the

known answer type and the set of keywords, the parser will form a SQL query and get the corresponding data from the database.

- For the user interface part, I use Django to build this website chat window. This chat window includes a simple chat feature which allow users to ask questions and get answer from the system.



Remaining work:

- For natural language generation part, I have implemented a simple NLG system to generate many single sentences. The problem of this system is that it cannot recognise other grammar rules and it is not flexible. Therefore, I am working on a new NLG system.
- Another work is to put the database on a cloud server.