

Kevin J. Shannon  
06 May 2011

A Thesis Presentation delivered in partial fulfillment of the designation University Honors  
Computer Science Senior Presentation Day, Cedar Falls, IA

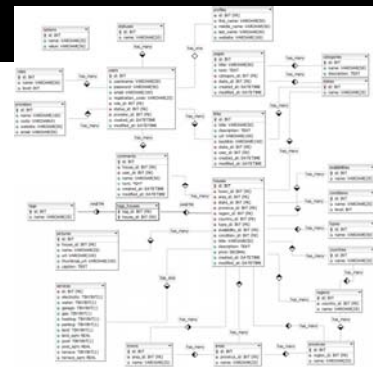
## Implementing a Natural Language to Structured Query Language Translator

### Definitions

- **Implementing a** – to build an actual system not just theory
- **Natural Language to** – human language, English
- **Structured Query Language** – SQL, computer language to get data from a database and database table, very popular and ubiquitous
- **Translator** – to make a one-to-one swap for a specific word or phrase

### The Problem

- **General:** How to get data in the hands of the people who can use it?
  - Completing reports
  - Decision making
  - Curiosity
- **Practical:** How to answer one-off questions?
- **Theoretical:** How to make humans and computers speak a common language?



[http://www.h3rald.com/files/italyisimply\\_database-architecture.png](http://www.h3rald.com/files/italyisimply_database-architecture.png)

### How does IT help?

- Understanding user needs by asking the right questions
- Creating and Designing new solutions to problems
- Delivering and following up
- Maintaining existing systems
- Expanding based on user needs

### General Solution

- Automate this process!
- Use an IT workflow to create reports
  - Stored Procedures
- Data Visualization
- GUI, like Business Objects or Crystal Reports
- Translator + encoding domain knowledge

## GUI, like BO or Crystal Reports

## Examples at UNI

- <https://www.ir.uni.edu/SAKPI/index.cfm?d=11&c=2&k=16>
- (CatID required)

## Challenges

- What domain do you pick?
- Web browser → Web server → Java → Database
- How do you encode the database information while still making it easy to edit/update?

## My Solution

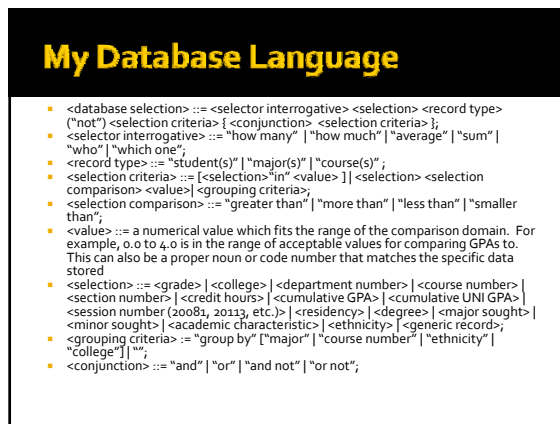
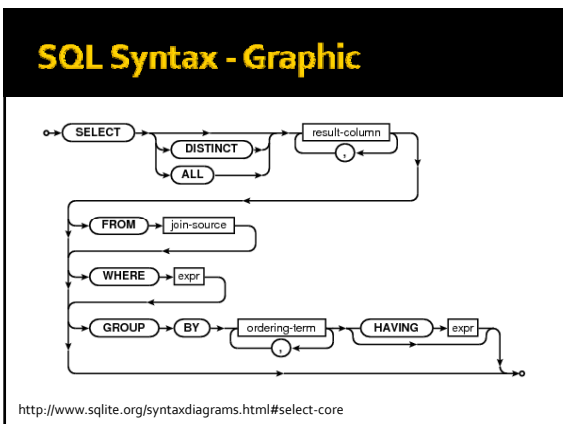
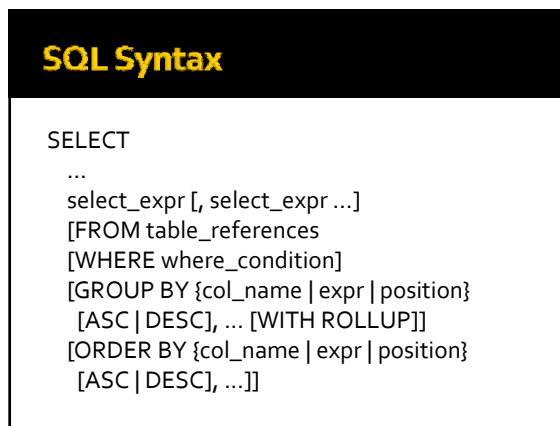
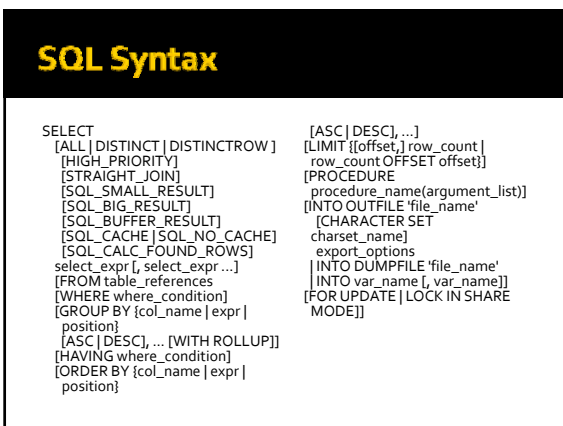
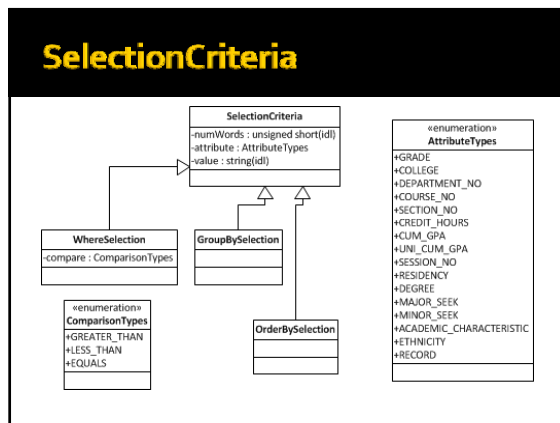
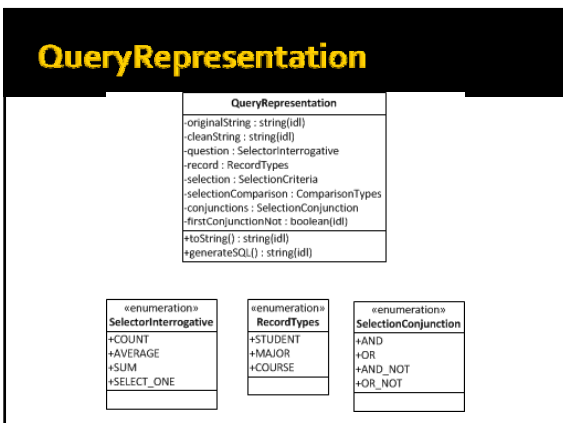
- Create a custom language that is a subset of English
- Translate word for word and phrase for phrase
- Use the translated parts to build an SQL statement
- Future Work: connect this directly to a database and have the system submit SQL

## Development Process

- Generate a language specification first
- Code to match first language specification
- Add a feature to the specification
  - Code this new feature
- Bugfix for screenshots for Thesis Reflection document and today's presentation

## My Solution

- Three Parts:
  1. Parser (NLPParserRunner.java & StringCleaner.java)
    - Splits up query into words, cleans up input
  2. Translator/Dictionary (NLPParser.java & ProperNounDatabase.java)
    - Looks for specific words/phrases
    - Adds "flags"/data to query representation for future interpretation
  3. Query Representation (QueryRepresentation.java)
    - Contains a method which generates SQL



## How does my language match SQL?

### SELECT

```

...
select_expr [, select_expr ...]
[FROM table_references]
[WHERE where_condition]
[GROUP BY {col_name | expr |
position}
[ASC | DESC], ... [WITH
ROLLUP]]
[ORDER BY {col_name | expr |
position}
[ASC | DESC], ...]]
    
```

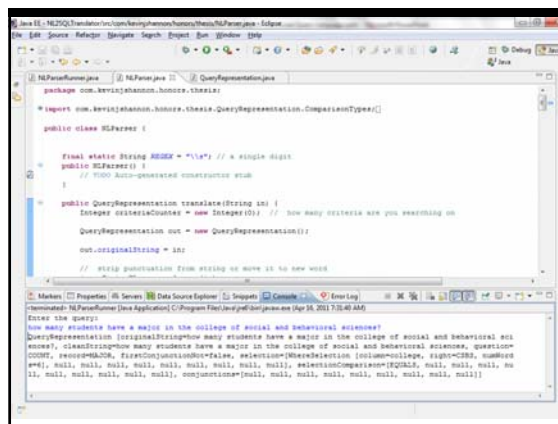
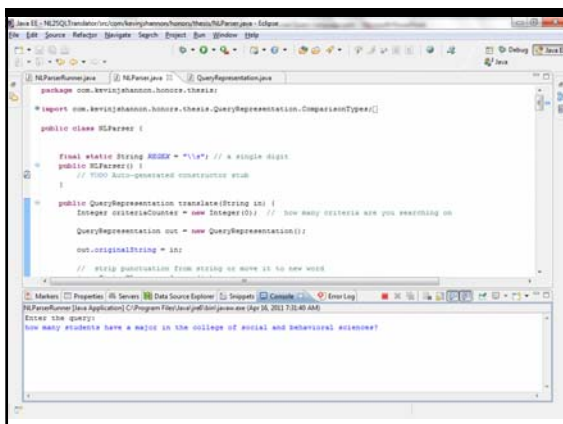
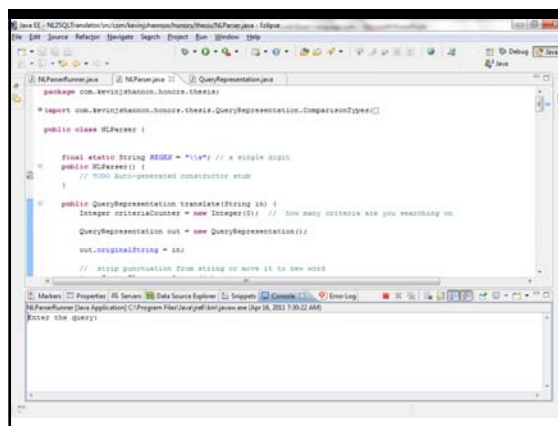
### MY DATABASE LANGUAGE

- <selector interrogative> := "how many" | "how much" | "average" | "sum" | "who" | "which one";
- <record type> := "student(s)" | "major(s)" | "course(s)";
- <selection comparison> := "greater than" | "more than" | "less than" | "smaller than";
- <selection> := <grade> | <college> | <department number> | <course number> | <section number> | <credit hours> | <cumulative GPA> | <cumulative UNI GPA> | <session number (2008, 2013, etc.)> | <residency> | <degree> | <major sought> | <minor sought> | <academic characteristic> | <ethnicity> | <generic record>;
- <grouping criteria> := "group by" | "major" | "course number" | "ethnicity" | "college" | ";

## Demo

## Examples

- How many student records are in the College of Social and Behavioral Sciences?
- How many major records are there in the graduate college
- how many student records are not in the college of business administration and not in college of humanities and fine arts group by major?
- How many student records have a gpa greater than 3.50?
- How many student records are in the Honors program and have a gpa greater than 3.50?
- what is the average uni gpa for students in the graduate college and uni gpa is greater than 3.5



```

package com.keriva.johnson.honora.thesis

import com.keriva.johnson.honora.thesis.QueryRepresentation.ComparisonTypes()

public class HParser {

    final static String REGEX = "\\d+" // a single digit
    public HParser() {
        // TODO Auto-generated constructor stub
    }

    public QueryRepresentation translate(String in) {
        Integer studentsCounted = new Integer(0) // how many criteria are you searching on
        QueryRepresentation out = new QueryRepresentation();
        out.originalString = in;

        // strip punctuation from string or move it to new word
    }
}

```

```

package com.keriva.johnson.honora.thesis

import com.keriva.johnson.honora.thesis.QueryRepresentation.ComparisonTypes()

public class HParser {

    final static String REGEX = "\\d+" // a single digit
    public HParser() {
        // TODO Auto-generated constructor stub
    }

    public QueryRepresentation translate(String in) {
        Integer studentsCounted = new Integer(0) // how many criteria are you searching on
        QueryRepresentation out = new QueryRepresentation();
        out.originalString = in;

        // strip punctuation from string or move it to new word
    }
}

```

```

package com.keriva.johnson.honora.thesis

import com.keriva.johnson.honora.thesis.QueryRepresentation.ComparisonTypes()

public class HParser {

    final static String REGEX = "\\d+" // a single digit
    public HParser() {
        // TODO Auto-generated constructor stub
    }

    public QueryRepresentation translate(String in) {
        Integer studentsCounted = new Integer(0) // how many criteria are you searching on
        QueryRepresentation out = new QueryRepresentation();
        out.originalString = in;

        // strip punctuation from string or move it to new word
    }
}

```

## What is Watson?

- <http://www.youtube.com/watch?v=FC3lryWr4c8>

## Questions?