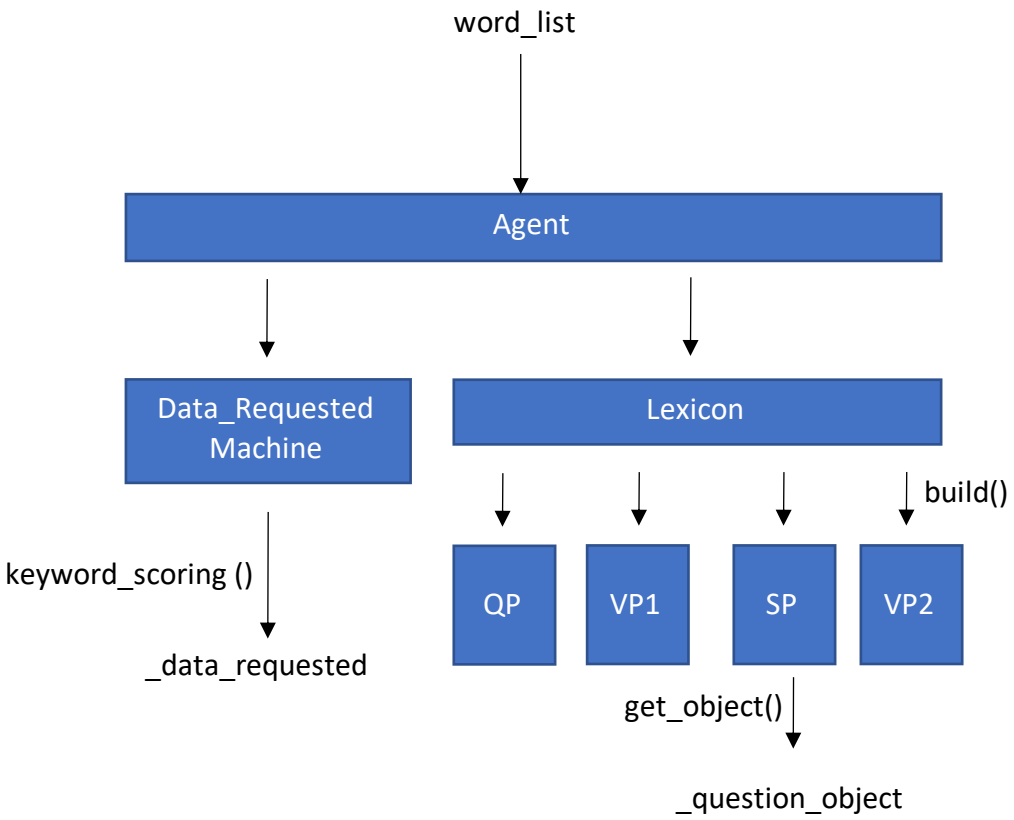
KBAI Project I

Overview / Block Diagram



In this agent, we tackle two problems. The first, and simpler, is establishing the type of data being requested by a question posed to our chatbot. For this, we use a simple voting method where the question is parsed by word. Every response type (DUEDATE, RELEASEDATE, WEIGHT, PROCESS, DURATION) has a unique list of single-word keywords as well as a list of paired (tuple) keywords. In our voting system, every single-word keyword is worth 1 point, and every tuple keyword is worth 5 points. Ties are broken by returning the topic with fewer possible points.

Our second dilemma, and the much more challenging one, is extracting the object the question is asking about. The domain of questions asked of this agent are questions beginning with Who, What, When, Where, Why, or How. These questions follow a very simple generic pattern.

Question Phrase + Verb Phrase 1 + Subject Phrase + Verb Phrase 2

Upon providing our Lexicon with the word list, it will go through the question and divide it into the 4 basic parts (if all exist). This process is begun by finding the first verb in the question. This verb will either be the only verb in the sentence, or it will be the auxiliary verb. Both of which occupy *Verb Phrase 1*. Everything before the first verb or auxiliary verb is considered a part of

the *Question Phrase*. Most of the time this is only the question word itself, but some more complicated sentence structures will occupy more of this block. A second verb is then searched for. If one is found, everything afterwards, along with the verb itself, is placed into *Verb Phrase 2*. If the secondary verb exists, then everything between the auxiliary verb and the second verb belongs to the *Subject Phrase*. This is a simple frame-like structure that allows us to break the question into almost tree-like segmentations. Each segment of the sentence can be further broken down. However, in this assignment, we were only required to extract the object of the question, which will always reside in the *Subject Phrase*.

At this stage, we must break the *Subject Phrase* down further. From this block, certain words are filtered out to create a list of object candidates. Determiners, conjunctions, common verbs, prepositions, common adjectives and possesives are all filtered out. Each object candidate is created with a frame. Each of these object candidate frames keeps track of the candidate itself, the index of the candidate in the word list, the proximity to a prior determiner, and the proximity to a prior preposition. The object was more likely to be directly following the determiner than a preposition, but it was not unusual to see the object following a preposition with no determiner. A hierarchy of cases is used to establish which candidate is actually the object. For objects not directly neighboring their nearest determiner or preposition, the neighboring words are attached as modifers in the case of compound word objects. For example, terms like 'first project' and 'project specification' will be returned together.

Our Agent was tested on a bank of 25 questions. Several are shown on the following page. Our agent scored 24/25 on _question_object and a 25/25 on _data_requested returns. The trickiest phrases for our agent to handle is when there are two strong candidates for the object. In these cases, it is common for the latter of the two to be the object, or for the one that is more closely preceeded by a determiner than a preposition. However, these are not always the case some questions will slip through the cracks.

The organization of my agent was mostly based on how I learned to build sentence diagrams in middle school. Going through that old process helped me to create the rules and relationships that began to govern how my Lexicon assorted the question into its building blocks. My agent's greatest strength is certainly the _data_requested. As the domain of topics for this agent is relatively small, a keyword voting function is efficient and appropriate.

```
{
    "question": "Where can I find the course syllabus",
    "object": "course syllabus",
    "data_requested": "PROCESS"
},
{
    "question": "Where is the first project located",
    "object": "first project",
    "data_requested": "PROCESS"
},
{
    "question": "Where can I find the course syllabus",
    "object": "course syllabus",
    "data_requested": "PROCESS"
},
{
    "question": "Where do I turn in project 1",
    "object": "project 1",
    "data_requested": "PROCESS"
},
{
    "question": "where is the project specification",
    "object": "project specification",
    "data_requested": "PROCESS"
},
{
    "question": "How much does the project contribute to my grade",
    "object": "project",
    "data_requested": "WEIGHT"
},
{
    "question": "How do I download project 1",
    "object": "project 1",
    "data_requested": "PROCESS"
},
{
    "question": "how much is the project worth",
    "object": "project",
    "data_requested": "WEIGHT"
},
{
    "question": "how long do we have to complete a project",
    "object": "project",
    "data_requested": "DURATION"
}
```