

PROMETHEUS AI

Phase 1

Sean Stappas
260639512

ECSE-498: Honours Thesis I

Supervised by: Prof. Joseph Vybihal

April 11, 2017

Abstract

Prometheus AI is a model of the human brain with the goal of controlling and coordinating multiple robots in different environments. This could be useful in environments hazardous for humans, such as in the aftermath of a nuclear power disaster, or in outer space. The model consists of four layers: the Neural Network (NN), the Knowledge Node Network (KNN), the Expert System (ES), and the Meta Reasoner (META). The NN classifies the signals coming from the robots' sensors and sends formatted tags to the KNN. The KNN represents memory and can initiate cascaded activation of memories in the form of tags, which are passed on to the ES. The ES is a simple logic reasoner and provides recommendations for actions to the META. The META represents high-level thinking and makes an intelligent decision for the robots. The assigned task this semester was to implement prototypes of the KNN and ES layers in Java. This was achieved using specific design criteria and extensive feedback from the project supervisor, Prof. Vybihal. Personal design criteria included using object-oriented programming principles, optimizing the system for speed and space efficiency, and maximizing code readability. Tests were created in TestNG and extensive documentation was written in Javadoc.

Acknowledgments

Elsa Riachi worked on the other two layers of the Prometheus AI (NN and META) in parallel with the work done in this report. She is also doing this project as part of an Honours Thesis. Many discussions were had together on the design of the system and how the layers should fit together.

Contents

List of Figures	3
List of Tables	3
1 Introduction	4
2 Background	4
2.1 Neural Network	4
2.2 Knowledge Node Network	4
2.3 Expert System	7
2.4 Meta Reasoner	7
2.5 Summary	8
3 Problem	8
4 Design Criteria	8
4.1 Efficiency	8
4.2 Object Oriented Design	8
4.2.1 Abstraction	9
4.2.2 Encapsulation	9
4.2.3 Polymorphism	9
4.2.4 Inheritance	9
5 Implementation	9
5.1 Tags	9
5.2 Knowledge Node Network	10
5.3 Expert System	11
5.4 Testing	12
5.5 Documentation	14
6 Plan for Next Semester	14
6.1 Finalization	14
6.1.1 Knowledge Node Network	14
6.1.2 Expert System	15
6.2 Integration	15
6.3 Testing	15
7 Impact on Society and the Environment	15
7.1 Use of Non-renewable Resources	15
7.2 Environmental Benefits	15
7.3 Safety and Risk	15
7.4 Benefits to Society	16
8 Conclusion	16
A UML Diagrams	17
References	19

List of Figures

1	Prometheus AI model.	4
2	High-level model of the Knowledge Node of the KNN.	5
3	Thinking forwards in the KNN.	6
4	Thinking backwards in the KNN.	6
5	Lambda thinking in the KNN.	6
6	Prometheus AI model with labeled input and output.	8
7	Setup for the <code>testKNN()</code> method.	13
8	Fall 2017 timeline.	14
A.1	UML diagram of the <code>tags</code> package.	17
A.2	UML diagram of the <code>knn</code> package.	17
A.3	UML diagram of the <code>es</code> package.	18
A.4	UML diagram of the <code>test</code> package.	18

List of Tables

1	Test setup for <code>testES()</code>	13
2	Test setup and activation for the ES portion of <code>testKNNandES()</code>	13

Listings

1	Method to think forwards until natural quiescence in the KNN.	10
2	Method to think forwards for a fixed number of cycles in the KNN.	10
3	Method to think forwards for a single cycle in the KNN.	11
4	Method to excite a KN.	11
5	Method to fire a KN.	11
6	Method to think until natural quiescence in the ES.	11
7	Method to think for a fixed number of cycles in the ES.	12
8	Method to think for a single cycle in the ES.	12
9	Method to add a Tag to the ES.	12

Abbreviations

The following abbreviations will be used throughout the report:

NN	Neural Network.
KNN	Knowledge Node Network.
KN	Knowledge Node.
ES	Expert System.
META	Meta Reasoner.
OOP	Object-Oriented Programming.

1 Introduction

The goal of this project is to create an artificial intelligence system called Prometheus to control and coordinate multiple robots. The robots themselves are expected to be relatively simple and do not make any decisions themselves. All the thinking and decision-making is done by Prometheus, which collects data from all the robots' sensors.

Applications for this type of system include robots in hazardous environments, such as in outer space (Mars, Moon, etc.), in nuclear plants after a nuclear disaster, and in military zones. The system could ideally learn from this data in a given environment and apply its learning to new environments.

The structure of the system is inspired from the functionality of the human brain, and is composed of the following four layers (in order of increasing abstraction): the Neural Network (NN), the Knowledge Node Network (KNN), the Expert System (ES), and the Meta Reasoner (META) [1]. The system can be seen in Figure 1 and will be described in Section 2.

2 Background

2.1 Neural Network

The NN layer consists of a network of neurons with a structure similar to neurons in the human

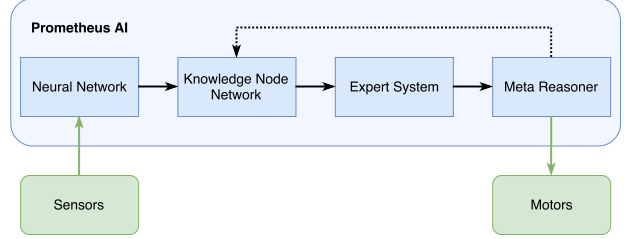


Figure 1: Prometheus AI model.

brain. It is the interface between the robots' sensors and the rest of the AI system.

The robots in Prof. Vybihal's lab have two types of sensors: cameras and ultrasonic sensors. The ultrasonic sensors can measure distance between the robot and nearby objects, and the camera can take images of what the robot is facing.

The NN gathers raw sensor data and will build an abstract view of the robot's surrounding environment. For each camera image, it will achieve two main goals:

1. Classify objects observed in the image.
2. Localize objects in the image.

To achieve these goals, a convolutional neural network will be used, which can determine which regions of the image are important and classify objects in those regions [2]. A 3D view of the world will also be generated. Using ultrasonic sensor readings and coordinate transformations from the world to the camera images, the observed objects can be localized in space.

Ultimately, the classification and localization of objects will produce abstract informational tags, which will be passed on to the KNN. These tags can therefore be measures of the position of an object, or other characteristics about an object.

2.2 Knowledge Node Network

The KNN layer represents memory in the human brain. It takes in the tags provided by the NN and outputs tags based on its knowledge of the environment. The KNN is based around interconnected Knowledge Nodes (KNs), which are

abstract structures representing memories and their connections to other memories. A simple model of a Knowledge Node (KN) can be seen in Figure 2.

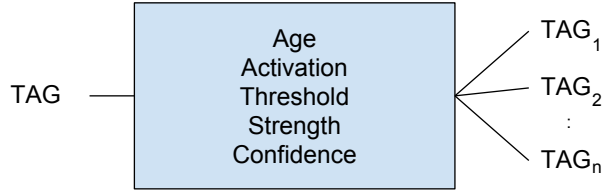


Figure 2: High-level model of the Knowledge Node (KN) of the KNN [3].

KNs have an input tag representing some information, and output tags representing information related to the input tag. Tags in the KNN can become “active”, meaning that they are observed or seen as true. For example, if the NN determines that it observes a ball and passes that information on to the KNN, then a “ball” tag may become active in the KNN. This would excite the KN with input tag “ball”. If that KN fires, other tags related to that observation can become active. For example, tags representing ball characteristics such as “round object” may become active. This tag in turn may be connected to another KN, potentially causing more activation.

There are different ways that KN excitation could be implemented. With simple linear activation, excitation would increment the node’s activation parameter, which initially starts at 0. If $activation \geq threshold$, the KN fires, causing the activation of the output tags. This description corresponds to forwards thinking. The activation can also be achieved using a sigmoid function, which is more representative of neurons in the brain [4].

There is a strength value associated with every KN which represents how much weight an activation has and therefore how quickly that node will fire. A simple implementation of strength would be as a constant coefficient multiplying the activation parameter. So, instead of checking when the activation is greater than the threshold, one would check if $activation * strength \geq threshold$, where the strength is positive. The

strength value could be 0 however, which effectively shuts off the KN. Another way of implementing strength is to check $activation + strength \geq threshold$. In this case, strength can be negative and hinder the firing of the node.

Strength can be seen as the firing predisposition of neurons in the brain as a result of learning [3]. Indeed, learning can increase the synaptic strength between neurons and cause early or late firing of those neurons [5]. For example, a person who has had a bad experience with spiders would fire their fear response upon seeing a spider more quickly than one without that fear.

Every KN can have a confidence value associated with it, representing how certain the KNN is that the input tag is true. In this case, the confidence value would be stored within the KN itself. These confidence values can come from the NN, since a neural network can have a confidence value when classifying objects. When initiating a think cycle, the confidence value at each stage can be multiplied with each other to produce a new confidence value, making the KNN less certain of a memory as it searches through its tree of KNs. This represents how belief changes when thinking. Indeed, some memories in the human brain require a great deal of thinking to reach and, as such, can be less certain than other memories. This can lead to the recollection of false memories [6].

Age represents how long it has been since a KN has been excited. The idea is that, after a KN has aged a certain amount of time, that node will be discarded, similarly to how old memories are discarded in the brain [7]. The intuitive implementation of aging is to constantly increment some age value associated with every KN, and discard nodes whose age is greater than some threshold. The age would be reset when the node is excited. This would however be unnecessarily computationally intensive, requiring some kind of constant updating of every single node. A more efficient way of doing it would be to save a timestamp for every KN when it is excited, and if the KNN attempts to excite a KN whose previous timestamp value is too far into the past, that node is instead deleted.

Thinking in the KNN represents an activa-

tion routine, which activates tags and potentially fires KNs. The KNN has three main ways of thinking: forwards, backwards, and lambda. The version of thinking to be done by the KNN is chosen by the META.

Forwards thinking is the simplest and is depicted in Figure 3. Firing a KN can cause forward activation of more KNs, hence the “forwards” naming.

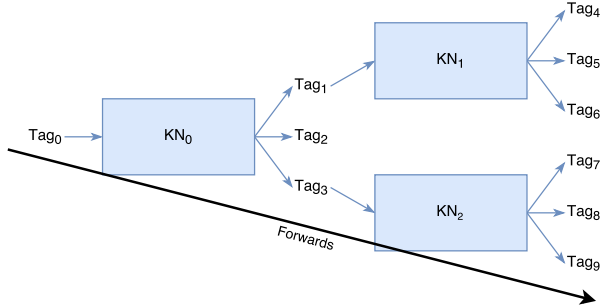


Figure 3: Thinking forwards in the KNN.

Thinking backwards starts at the output tags of KNs, and works backwards, as can be seen in Figure 4. In its simplest form, it checks output tags of KNs and, if all of them are active, the input tag must be active as well, so that KN is fired.

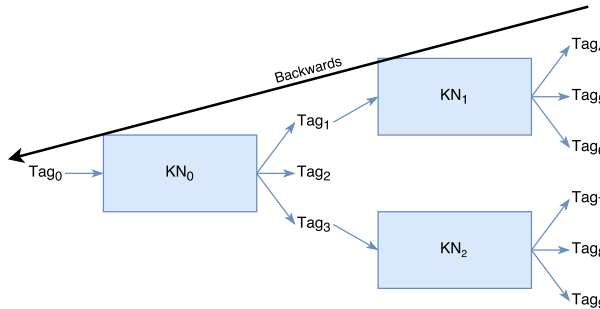


Figure 4: Thinking backwards in the KNN.

This type of thinking can be more difficult to implement than forwards thinking, since one must decide what percentage of active output tags of a KN corresponds to an active input tag. For instance, should a KN fire when 100% of its output nodes are active, or when 75% are active? This relates to the confidence with which the KNN believes the tag associated with that

node to be true. As a concrete example, if you observe an object that has four wheels, seats, and a steering wheel, how confident are you that that object is a car? Realistically, humans with often classify what they observe with some uncertainty [8] and this is what can be represented with backwards thinking. This type of thinking occurs constantly in the background in humans [3], and this is something to keep in mind during implementation.

Lambda thinking uses a combination of forwards and backwards, and can be seen in Figure 5. It is called “lambda” thinking because the shape of the thinking trajectory matches a Greek uppercase lambda (Λ). It first looks at output tags of KNs and propagates activation backwards, similar to backwards thinking. After a certain amount of nodes are fired, it will then start activating forwards, similar to thinking forwards. An interesting question is how far backwards should lambda thinking go before starting to cascade forwards? This relates to how general one wants to explore before searching for a more specific value.

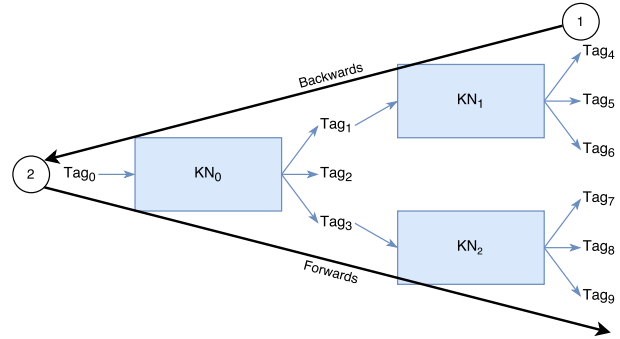


Figure 5: Lambda thinking in the KNN.

This type of thinking occurs in humans when using analogical reasoning to find a memory [9]. In essence, when a person wants to locate a memory that is not directly accessible, they will explore related memories, move their way “backwards” to concepts related to the desired memory, and think “forwards” to focus in on the desired memory. For example, if one is asked where they were in 2002 at a specific date, they most likely would not remember. If they explore related dates and events in their life, however, they

may be able to narrow down and extract that memory. In the context of the Prometheus system, lambda thinking will be attempted if all other forms of thinking fail (forwards and backwards).

All forms of thinking can continue until there are no more KNs to activate, which corresponds to natural quiescence. There can also be a fixed number of thinking cycles, which represents how much effort is being put into thinking. Indeed, in humans, thinking is done with varying degrees of effort [10], and it would be interesting to model this.

The result of all forms of thinking is a collection of activated tags, which are passed on to the ES.

2.3 Expert System

The ES layer is a basic logic reasoner. It is not aware of its current reality or any context. It takes in the tags provided by the KNN and interprets them as either facts, recommendations or rules.

Facts are simple calculus predicates showing that something is true. Here are some examples of fact tags and their meanings:

- (A) A itself is true or active.
- ($A = 1$) A is equal to 1.
- ($A > 1$) A is greater than 1.
- (A ?) A can take any value.

As a more concrete example, a fact can represent a certain measurement, like (*distance = 5*) representing the robot's distance from a wall as measured by one of its sensors.

Recommendations represent suggestions for actions to be taken by a robot. For example, (*#turn.left*) is a recommendation for a robot to turn left, if it sees a wall directly in front of it and must avoid it, for example. These are recommendations and not commands because the META can decide whether or not to actually take that action.

Rules are many-to-many structures with facts as inputs and recommendations or facts as

outputs. This can be seen in Equation (1), where $m \geq 1$ and $n \geq 1$, i.e., there must be at least one input tag and one output tag.

$$Fact_1 \cdots Fact_m \rightarrow Tag_1 \cdots Tag_n \quad (1)$$

When all the input tags become active, the output tags become active and the rule itself is also said to be active. In this way, a rule can represent a logical AND of all its input tags. The runtime of the ES consists of the following general steps [11]:

1. Reset.
2. Add facts and rules.
3. Think.
4. Send recommendations to META.

The most important part of the previous process is the thinking stage, which represents the activation routine of the ES. This consists of first iterating through all the rules in the ES and checking if they are active by inspecting the lists of facts and recommendations. Rules may then become active and cause cascading activation of more rules. This can continue until there are no more rules to activate, which corresponds to natural quiescence. There can also be a fixed number of thinking cycles, which represents the effort put into thinking, similarly to the KNN.

The recommendations activated as a result of thinking are passed on to the final layer, the META.

2.4 Meta Reasoner

The META layer represents high-level reasoning in human brains. It is aware of its environment and context, and makes decisions based on what it believes to be right. It is paranoid, and constantly checks whether the tags reported by the rest of the AI system make sense based on its expected view of the world. If it decides to make a decision, it sends a command to the actuators of the robots to decide how to move. If it is not happy with the recommendation(s) from the ES, it may initiate another think cycle in the KNN to generate new recommendations(s).

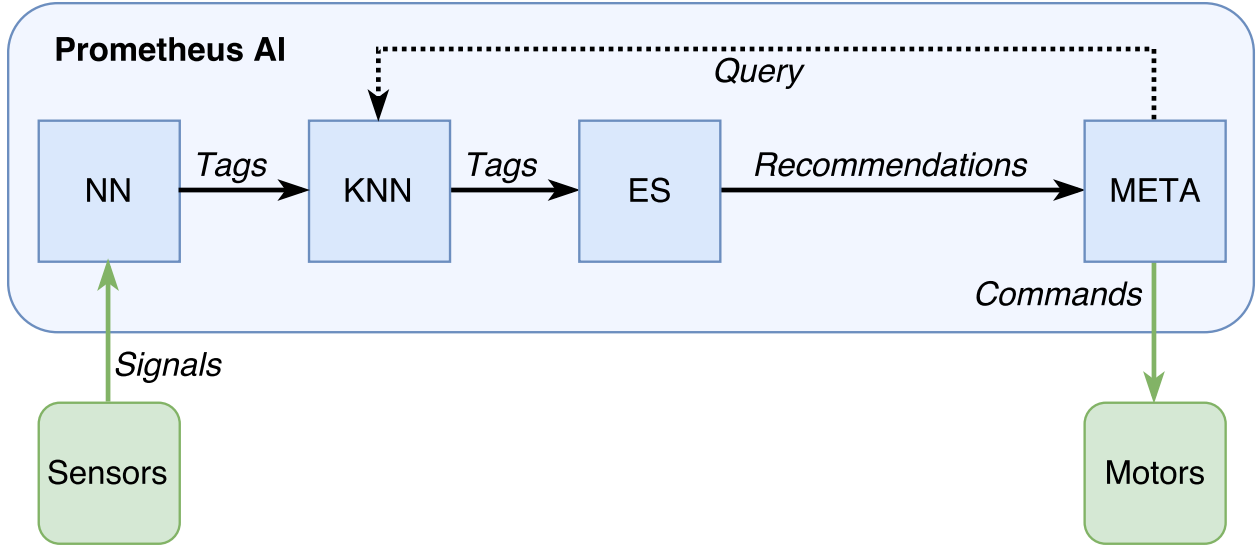


Figure 6: Prometheus AI model with labeled input and output.

2.5 Summary

With this full description of the Prometheus AI model, the system with labeled input and output can be seen in Figure 6.

3 Problem

The assigned task was to construct two out of the four layers listed in Section 2: the ES and the KNN. The other two layers are to be completed by another Honours Thesis student.

As a deliverable for the end of this first semester, it was required to complete a prototype of these two layers, with basic versions of the functionality described in Section 2. This was to be done entirely in Java.

4 Design Criteria

4.1 Efficiency

A very important consideration when designing the system is speed. Since the robots may have to react very quickly to stimuli in the environment, the reasoning in the AI must be as fast as possible. This is especially true in the hazardous environments for which this system could be useful for, as specified in Section 1. An example of a design choice that was made

to improve speed is the use of Java `Sets` for most of the collections in the ES and KNN layers. The original specifications mentioned using `ArrayLists`, but, since there is no specific iteration order necessary for most operations in the ES and KNN, these collections were changed to `HashSets`. This is faster because `HashSets` have $O(1)$ access time, whereas `ArrayLists` have $O(N)$ access time, where N is the number of elements in the collection. `HashSets` also have the advantage of only permitting unique elements.

4.2 Object Oriented Design

Another important choice is to leverage object-oriented design as much as possible. Object-oriented programming (OOP) allows extensive planning before even beginning to write code, which can identify any flaws in the initial design. It also allows the code to be very clean and reusable [12]. Since Java is the programming language chosen for the project, OOP is also the natural way to proceed. To follow OOP, the ES and KNN will be designed around Java classes and the methods associated with those classes. OOP principles such as polymorphism and encapsulation will be followed closely.

4.2.1 Abstraction

One important design aspect is that the system should be as abstract as possible, while still performing its desired task. For instance, the system should be general enough to perform under simulations, as well as in real-life environments. It should also ideally be able to perform in vastly different environments, with different tasks.

One example of making use of abstraction is the choice to create a **PrometheusLayer** interface for both the ES and KNN, since they share some functionality, like the ability to think.

4.2.2 Encapsulation

Each layer of the system has unique, localized functionality that does not need to be visible from the rest of the system. For instance, the intricacies of the **think()** method in the ES and KNN layers do not need to be known to the rest of the system.

4.2.3 Polymorphism

Many methods in the system can have different functionality depending on context. For instance, the **think()** method in the KNN and ES layers can take an optional number of think cycles to specify a threshold for quiescence.

4.2.4 Inheritance

Careful thought was put into how objects may inherit functionality from each other. This can most clearly be seen with the implementation of the tags, which will be described in Section 5.

5 Implementation

The details of how the prototype of the system was implemented in Java will now be discussed.

5.1 Tags

As described in Section 2, the entire system revolves around tags passed from layer to layer. For this reason, a lot of thought was put into the proper design of these tags.

The tags need to be as general as possible. A natural choice for this structure would be a Java **String**, which would be relatively simple to pass around the system. However, these tags represent various concepts; each tag can either be a fact, a recommendation, or a rule. If implemented as **Strings**, the tags would have to be encoded on creation to represent each concept and decoded on use to retrieve the important information. This seems like a bad use of the OOP principles of Java. Furthermore, if specific functionality is needed in the future for each tag type, that can easily be implemented with a Java class. For these reasons, the tags are implemented using a **Tag** Java class, with **Recommendation**, **Fact**, and **Rule** subclasses. This should also make manipulating the Tags faster, while incurring a slight memory overhead. To store these Tags in a database, they can be converted to JSON format. On read from the database, they can be easily decoded.

The **Tag** class has an associated **Type**, which can take the following **enum** values: **FACT**, **RECOMMENDATION**, and **RULE**. This is used to distinguish the three types of **Tags**.

The **Fact** and **Recommendation** classes at this point are little more than wrapper classes around a **String** value. More interesting features will be added later on (see Section 6). The **Rule** class has the following important fields:

inputFacts Array of input **Facts**.

outputTags Array of output **Tags**.

The **Tag** class itself was made abstract. This means that an object may not be directly instantiated as a **Tag**, but must be instantiated as one of its subclasses. This makes sense, since a tag *must* be one of the three types: rule, recommendation or fact.

All of these classes were placed inside the **tags** package. A UML diagram of the **tags** package can be seen in Figure A.1 of the Appendix.

5.2 Knowledge Node Network

All code relating directly to the KNN was placed in the `knn` package of the project. A UML diagram of the `knn` package can be seen in Figure A.2 of the Appendix.

The KNN layer is based around the `KnowledgeNodeNetwork` Java class. This class has the following fields:

`mapKN` One-to-one `HashMap` of input `Tags` to associated `KnowledgeNodes`.

`activeTags` `HashSet` of active `Tags`, corresponding to input `Tags` of fired `KnowledgeNodes`.

The `KnowledgeNode` class implements the functionality of a `KN`, which has the following important fields to implement the functionality described in Section 2:

`inputTag` Input `Tag`.

`outputTags` Array of output `Tags`.

`activation` `int` starting at 0, incrementing when the `KnowledgeNode` is excited.

`threshold` `int` threshold that causes firing of the `KnowledgeNode`.

`strength` `int` that biases the activation of a `KnowledgeNode`, causing early firing. The `KnowledgeNode` is fired if $(\text{activation} \cdot \text{strength}) \geq \text{threshold}$.

`confidence` `int` representing the belief that the `inputTag` is true (0 to 100).

`age` `int` representing the age of the `KnowledgeNode`.

The most important method in the `KnowledgeNodeNetwork` is `think()`, which chooses either `thinkForwards()`, `thinkBackwards()`, or `thinkLambda()`. These methods implement the functionality described in Section 2 and return the `Tags` activated as

a result of thinking. At this point, `think()` chooses `thinkForwards()` by default. In the future, the choice will be determined by `META` (see Section 6).

Without a parameter, the `think()` method runs to natural quiescence. There is also an overloaded version of `think()` that takes an `int numberOfCycles` as a parameter. This parameter represents the thinking effort described earlier. Similarly, `thinkForwards()`, `thinkBackwards()`, and `thinkLambda()` all have overloaded versions with `numberOfCycles` as a parameter.

Listing 1: Method to think forwards until natural quiescence in the KNN.

```
1 private Set<Tag> thinkForwards() {
2     Set<Tag> totalActivatedTags = new
        HashSet<>();
3     Set<Tag> activatedTags;
4     do {
5         activatedTags = forwardThinkCycle();
6         totalActivatedTags.addAll(activatedTags);
7     } while (!activatedTags.isEmpty());
8     return totalActivatedTags;
9 }
```

Listing 2: Method to think forwards for a fixed number of cycles in the KNN.

```
1 private Set<Tag> thinkForwards(int
    numberOfCycles) {
2     Set<Tag> totalActivatedTags = new
        HashSet<>();
3     for (int i = 0; i < numberOfCycles; i++) {
4         Set<Tag> activatedTags =
            forwardThinkCycle();
5         if (activatedTags.isEmpty()) {
6             break;
7         }
8         totalActivatedTags.addAll(activatedTags);
9     }
10    return totalActivatedTags;
11 }
```

Currently, only the `thinkForwards()` method is fully completed. The version that runs until natural quiescence can be seen in Listing 1. The version with a fixed number of thinking cycles can be seen in Listing 2. One can see that both `thinkForwards()` methods call `forwardThinkCycle()`, which can be seen

in Listing 3. This is where Tags can become active (line 9), being added into `activeTags`.

Listing 3: Method to think forwards for a single cycle in the KNN.

```

1 private Set<Tag> forwardThinkCycle() {
2     Set<Tag> allPendingTags = new HashSet<>();
3     for (Tag tag : activeTags) {
4         if (mapKN.containsKey(tag)) {
5             Set<Tag> pendingTags =
6                 excite(mapKN.get(tag));
7             allPendingTags.addAll(pendingTags);
8         }
9     }
10    activeTags.addAll(allPendingTags);
11    return allPendingTags;
12 }
```

One can see that the `forwardThinkCycle()` method calls `excite()` (line 5) on a `KnowledgeNode`, which returns the Tags activated as a result of excitation. The method can be seen in Listing 4. This, in turn, may call `fire()` on line 5 to fire the `KnowledgeNode`, activate its output tags, and return the Tags activated as a result of firing. The method can be seen in Listing 5.

Listing 4: Method to excite a KN.

```

1 private Set<Tag> excite(KnowledgeNode kn) {
2     Set<Tag> pendingTags = new HashSet<>();
3     kn.activation++;
4     if (kn.activation * kn.strength >=
5         kn.threshold) {
6         pendingTags = fire(kn);
7     }
8     return pendingTags;
9 }
```

Listing 5: Method to fire a KN.

```

1 private Set<Tag> fire(KnowledgeNode kn) {
2     Set<Tag> pendingTags = new HashSet<>();
3     for (Tag tag : kn.outputTags) {
4         if (!activeTags.contains(tag)) {
5             pendingTags.add(tag);
6         }
7     }
8     return pendingTags;
9 }
```

A simple version of `thinkBackwards()` was

also completed, activating a KN only if all its output Tags are active.

5.3 Expert System

All code relating directly to the ES was placed in the `es` package of the project. A UML diagram of the `es` package can be seen in Figure A.3 of the Appendix.

The ES layer is based around the `ExpertSystem` Java class. This class has the following fields:

<code>readyRules</code>	<code>HashSet</code> of Rules that have not been activated yet.
<code>activeRules</code>	<code>HashSet</code> of active Rules.
<code>facts</code>	<code>HashSet</code> of active Facts.
<code>recommendations</code>	<code>HashSet</code> of active Recommendations.

The most important method in the `ExpertSystem` is `think()`, which implements the functionality described in Section 2. Without a parameter, the `think()` method runs to natural quiescence. This can be seen in Listing 6.

Listing 6: Method to think until natural quiescence in the ES.

```

1 public Set<Tag> think() {
2     Set<Tag> allActivatedTags = new HashSet<>();
3     Set<Tag> activatedTags;
4     do {
5         activatedTags = thinkCycle();
6         allActivatedTags.addAll(activatedTags);
7     } while (!activatedTags.isEmpty());
8     Set<Tag> activatedRecommendations = new
9         HashSet<>();
10    for (Tag tag : allActivatedTags) {
11        if (tag.isRecommendation())
12            activatedRecommendations.add(tag);
13    }
14    return activatedRecommendations;
15 }
```

There is also an overloaded version of `think()` that functions similarly to that of the `KnowledgeNodeNetwork`. This can be seen in Listing 7. Both variants of `think()` return the

Set of Recommendations that become active as a result of thinking, which is passed on to the META.

Listing 7: Method to think for a fixed number of cycles in the ES.

```

1 public Set<Tag> think(int numberOfCycles) {
2     Set<Tag> allActivatedTags = new HashSet<>();
3     for (int i = 0; i < numberOfCycles; i++) {
4         Set<Tag> activatedTags = thinkCycle();
5         if (activatedTags.isEmpty())
6             break;
7         allActivatedTags.addAll(activatedTags);
8     }
9     Set<Tag> activatedRecommendations = new
10        HashSet<>();
11     for (Tag tag : allActivatedTags) {
12         if (tag.isRecommendation())
13             activatedRecommendations.add(tag);
14     }
15     return activatedRecommendations;
16 }

```

Listing 8: Method to think for a single cycle in the ES.

```

1 private Set<Tag> thinkCycle() {
2     Set<Tag> activatedTags = new HashSet<>();
3     Set<Rule> pendingActivatedRules = new
4        HashSet<>();
5     for (Rule rule : readyRules) {
6         boolean shouldActivate = true;
7         for (Fact fact : rule.inputFacts) {
8             if (!facts.contains(fact)) {
9                 shouldActivate = false;
10                break;
11            }
12        }
13        if (shouldActivate)
14            pendingActivatedRules.add(rule);
15    }
16    for (Rule rule : pendingActivatedRules) {
17        readyRules.remove(rule);
18        activeRules.add(rule);
19        for (Tag tag : rule.outputTags) {
20            if (!facts.contains(tag) &&
21                !recommendations.contains(tag)) {
22                activatedTags.add(tag);
23                addTag(tag);
24            }
25        }
26    }
27    return activatedTags;
28 }

```

Both versions of `think()` call a method named `thinkCycle()` representing a single cycle of thinking. The code for this method can be

seen in Listing 8.

Listing 9: Method to add a Tag to the ES.

```

1 public boolean addTag(Tag tag) {
2     switch (tag.type) {
3         case RULE:
4             return addRule((Rule) tag);
5         case FACT:
6             return addFact((Fact) tag);
7         case RECOMMENDATION:
8             return
9                 addRecommendation((Recommendation)
10                    tag);
11     }
12     return false;
13 }

```

Finally, we see in the `thinkCycle()` method a call to the `addTag()` method on line 21, which adds a `Tag` to the ES depending on its type. The code can be seen in Listing 9. The method is `public` because it could be used by a user of the ES to set the initial data structures. It returns `true` if the `Tag` is added successfully. The `addRule()`, `addFact()`, and `addRecommendation()` methods simply add a `Tag` to the `readyRules`, `facts`, or `recommendations` Sets, respectively, and return `true` if the ES did not already contain the `Tag` to be added.

5.4 Testing

All tests on the system were conducted using the TestNG framework in Java, which provides a simple and intuitive way to create assertions in tests. These tests were placed in the `test` package of the project. A UML diagram of the `test` package can be seen in Figure A.4 of the Appendix. The primary methods of interest in the `test` package are `testKNN()`, `testES()`, and `testKNNandES()`.

The test setup for the `testKNN()` method can be seen in Figure 7. It creates a KNN, with every node having a `threshold` value of 1 to simplify the activation process. The system starts out with `A` as the only active `Tag`, and `testKNN()` will make the KNN `think()` until natural quiescence. It is therefore expected all the `Tags` shown in Figure 7 to become active at the end.

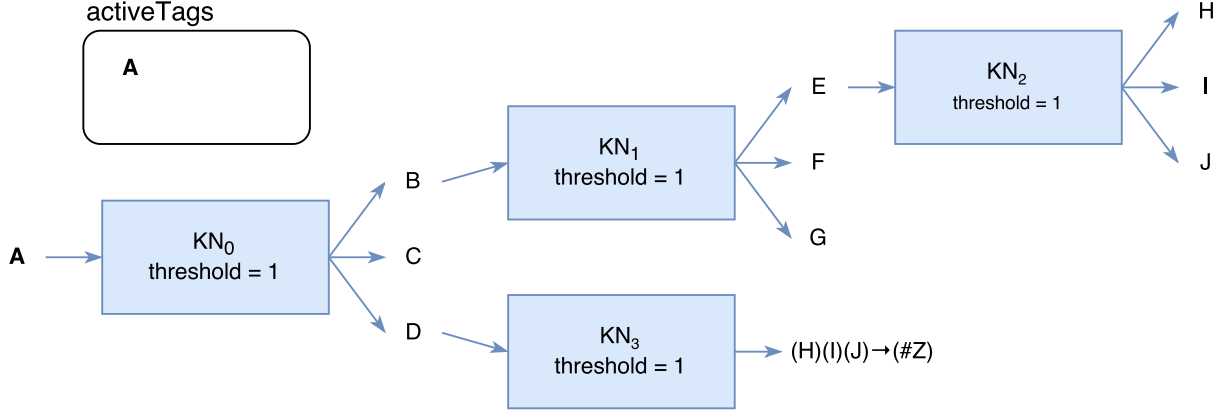


Figure 7: Initial setup for the *testKNN()* method.

Table 1: Test setup for *testES()*. Middle activation steps omitted.

State	Ready Rules	Active Rules	Active Facts	Active Recommendations
Initial	$(A)(B) \rightarrow (D)$ $(D)(B) \rightarrow (E)$ $(D)(E) \rightarrow (F)$ $(G)(A) \rightarrow (H)$ $(E)(F) \rightarrow (\#Z)$		$(A), (B)$	$(\#X), (\#Y)$
\vdots	\vdots	\vdots	\vdots	\vdots
Final	$(G)(A) \rightarrow (H)$	$(A)(B) \rightarrow (D)$ $(D)(B) \rightarrow (E)$ $(D)(E) \rightarrow (F)$ $(E)(F) \rightarrow (\#Z)$	$(A), (B),$ $(D), (E)$ (F)	$(\#X), (\#Y), (\#Z)$

Table 2: Test setup and activation for the ES portion of *testKNNandES()*.

State	Ready Rules	Active Rules	Active Facts	Active Recommendations
Initial	$(H)(I)(J) \rightarrow (\#Z)$		$(B), (C), (D)$ $(E), (F), (G)$ $(H), (I), (J)$	
Final		$(H)(I)(J) \rightarrow (\#Z)$	$(B), (C), (D)$ $(E), (F), (G)$ $(H), (I), (J)$	$(\#Z)$

The initial and final states are both asserted with TestNG in `testKNN()`, with positive results.

The test setup for the `testES()` method can be seen in Table 1. The columns from left to right correspond to the elements in the `readyRules`, the `activeRules`, the `facts`, and the `recommendations` respectively. The first row corresponds to the initial setup of the ES, and the last row corresponds to the expected final state. Both the initial and final states are asserted with TestNG in `testES()`, with positive results.

The `testKNNandES()` method has the same setup as `testKNN()` in Figure 7, except the output active Tags from the KNN are passed on to the ES. The resulting setup and activation in the ES can be seen in Table 2. The initial and final states were also tested with TestNG with positive results.

5.5 Documentation

Proper documentation of the source code is very important. This is to ensure that anyone wanting to work with the code which was designed has an easy time doing so. Also, for anyone wanting to understand how the system works, documentation is essential. This documentation was achieved through Javadoc¹ and UML diagrams (see Appendix A).

6 Plan for Next Semester

The plan for next semester is to finalize the two layers that were started this semester (KNN and ES), integrate these layers with the remaining two layers (NN and META), and to test the system in simulation and on the robots available in Prof. Vybihal's lab. The expected timeline can be seen in Figure 8.

6.1 Finalization

First, the KNN and ES will need to be finalized. It is expected that this will take up the first month of next semester.

¹The Javadoc can be found here: <http://cs.mcgill.ca/~sstapp/prometheus/index.html>

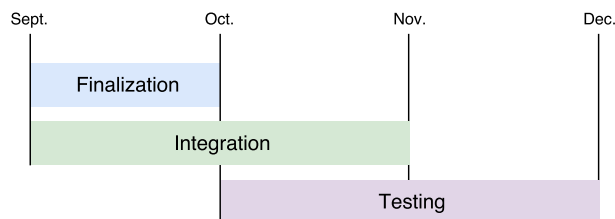


Figure 8: Fall 2017 timeline.

6.1.1 Knowledge Node Network

Many features of the KNN layer are still to be implemented, such as backwards and lambda thinking, confidence, and aging.

Backwards thinking can be implemented using a thread in Java, constantly running in the background at some rate.

Some thought still needs to be put into the specifics on implementation of lambda thinking. Most specifically, there will have to some scoring method to determine how far to search backwards before moving forwards.

Confidence will have to be implemented. Confidence values will probably be associated with every individual Tag, as well as in an absolute sense in the KNN. Activation through the layers of the KNN can theoretically then reduce the confidence that the output Tags are true, since at every step the confidence should be multiplied with the previous value.

Aging is also still left to be implemented, with a timestamp system as described in Section 2.

Activation of KNs with sigmoid functions will also be looked into. This can be easily implemented using a cache of known sigmoid values.

Also, more thought will have to be put into the design of the KNN to allow cyclic graphs, representing recursive memories in the human brain.

Finally, there will be more features to be added in the KNN which are not described in Section 2, since they are still in the process of being thought out. These features include learning, long-term memory and attention.

6.1.2 Expert System

Some features of the ES layer are still to be implemented, such as more complex **Fact** checking. Indeed, currently, the ES only checks for strict equality between **Facts**, but it would be interesting to implement checking for “greater than” or “less than” relations as well.

6.2 Integration

One very important task left to be done is to integrate the two layers described in this report (ES and KNN) with the other layers developed separately (NN and META). Ideally, the layers should be able to work together, but there will surely be some conflicts at the interface of the layers. These will have to be resolved when the time comes.

The interface between the NN and the KNN will have to be tested. There may have to be some conversion between what the NN will output and what the KNN expects (**Tags**). The META layer is also still to be implemented. It is crucial, since it will decide the way the KNN will think.

The integration section is expected to take around 2 months, with the the first month overlapping with the finalization section.

6.3 Testing

Finally, the system will have to be tested. This is expected to take around 2 months, with the first month overlapping with the last month of integration.

The first and easiest way to test would be in a simulated environment. One simulator that may be used is Simbad, which is a Java 3D robot simulator [13]. This can allow for some early debugging and fixes.

Once the simulation testing is completed and working properly, the system can be tested in the lab. Prof. Vybihal’s lab has multiple robots with ultrasonic sensors, and these will be the test subjects of this phase.

7 Impact on Society and the Environment

7.1 Use of Non-renewable Resources

As purely a software project, there are no physical materials needed to construct this system. For this reason, less emphasis will be put on this section. The only physical resources potentially needed are the materials needed to construct the robots and the computers to house the software. One area of concern could be the energy source of the robots themselves, which would probably vary depending on context. Ideally, the source should be a renewable one. For example, solar panels could be placed on the robots to provide energy.

7.2 Environmental Benefits

Prometheus could be used in a context beneficial to the environment. For example, the system could be used as a tool to control robots after oil spills, where the robots could theoretically contain the problem faster than humans and thus limit the risk on the environment. There is already research being done on employing robots in this context. Indeed, MIT’s Senseable City Lab has been working on Seaswarm, a system composed of a fleet of vehicles to help clean future oil spills [14]. An AI like Prometheus to control a swarm of robots like this could be very valuable.

7.3 Safety and Risk

It is critical that, once this system is completed, it is used in an ethical way, and for the right purposes. One example of use that may cause ethical concern is in a military setting, where an AI system like the one described here could be used in a battlefield in place of soldiers. Indeed, the US Department of Defense has plans to employ AI for autonomous weapons to attack targets without human intervention in the future [15]. This would have the advantage of potentially saving human soldiers’ lives [16]. However, there are also concerns because this would make it much easier to start a battle.

More than 3000 AI and robotics researchers have signed an open letter arguing against a military AI arms race [17]. They argue that autonomous weapons would not be beneficial to society, since they would be ideal for assassinations, destabilizing countries or selectively subduing or killing a population.

Another possible issue is the future loss of jobs to be done by humans, with automated systems like Prometheus replacing physical labor. This problem can be solved as a society, perhaps by having more support for universal basic income (UBI). Tesla CEO Elon Musk is a proponent of the idea, saying that UBI will be necessary in the future [18]. Bill Gates suggests possibly having a tax on robots to help pay for this income [19].

In the very long term, there are concerns with the possibility that an AI system might achieve intelligence and awareness close to a human. If such an AI were to obtain “consciousness” in its own way, should that entity be entitled to its own rights, like humans or animals are? The issue has been mentioned by the Institute for the Future, calling for the possibility of “robo-rights” in the future [20].

7.4 Benefits to Society

This type of system could be extremely useful in many contexts in society. For instance, the system could be used to send and control robots in an area that would otherwise be very dangerous for humans. For example, robots are often used in the aftermath of nuclear disasters to prevent the unnecessary loss of human life. Indeed, even in the Chernobyl nuclear disaster, the Soviet authorities employed the use of robots to avoid losses of human life [21]. With an AI like Prometheus, the robots could be controlled in an intelligent manner.

The system could also be used to further space exploration, with an AI controlling multiple robots exploring the surface of Mars, for instance. The value of AI in this context has been clearly shown, with the Mars Curiosity rover recently being upgraded to have its own AI system using computer vision to identify rocks [22].

This system can therefore help further important cutting-edge research in space exploration.

8 Conclusion

This semester, prototypes of the Expert System (ES) and Knowledge Node Network (KNN) layers of the Prometheus AI model were completed in Java based on personal design criteria and feedback from the supervisor. The design of the prototypes has required a great deal of thought and planning to ensure efficiency and proper behaviour. These prototypes were tested using the TestNG Java framework with positive results, and extensive Javadoc was produced.

The main goal for next semester is to finalize the entire system, implementing more complex features that were omitted for this prototype stage. This will also require proper integration between the work done in this report (KNN and ES), and outside this report (NN and META). Finally, the entire system will have to be tested, in simulation and with physical robots.

A UML Diagrams

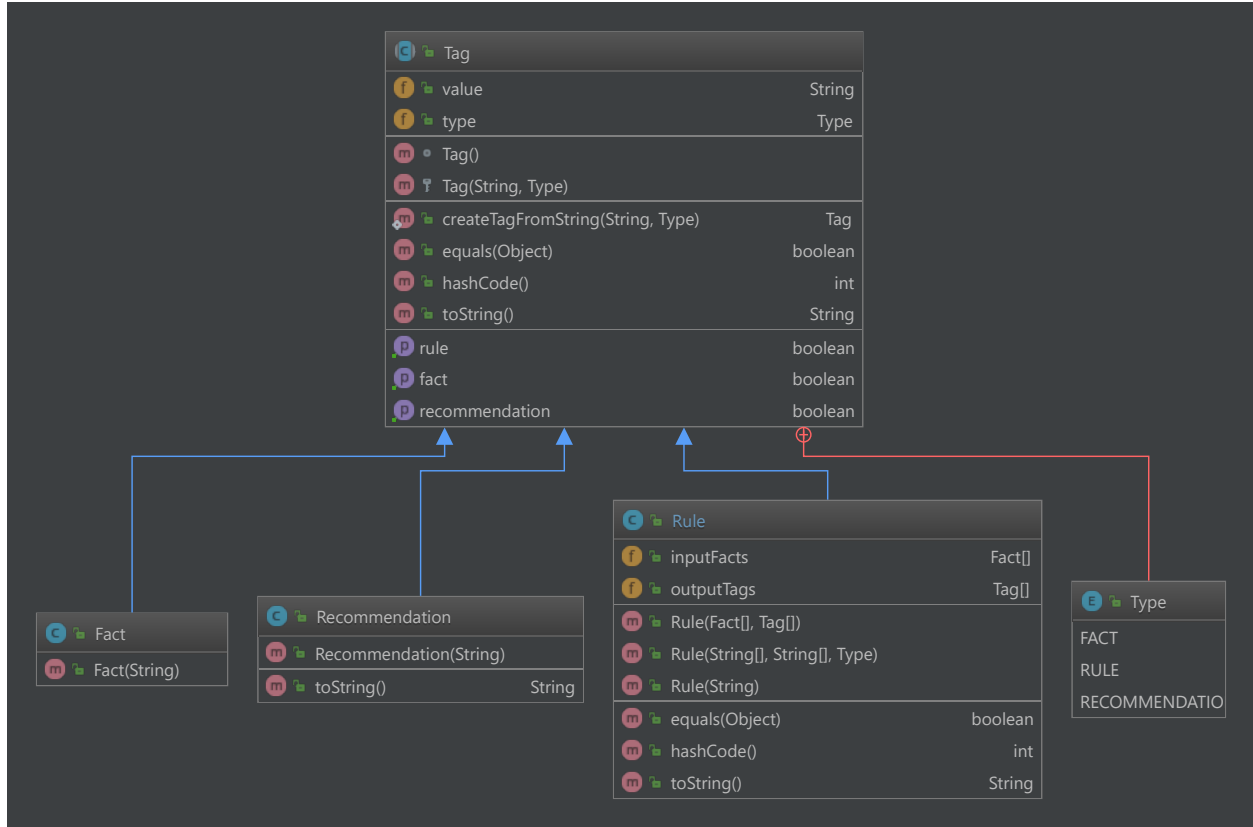


Figure A.1: UML diagram of the *tags* package.

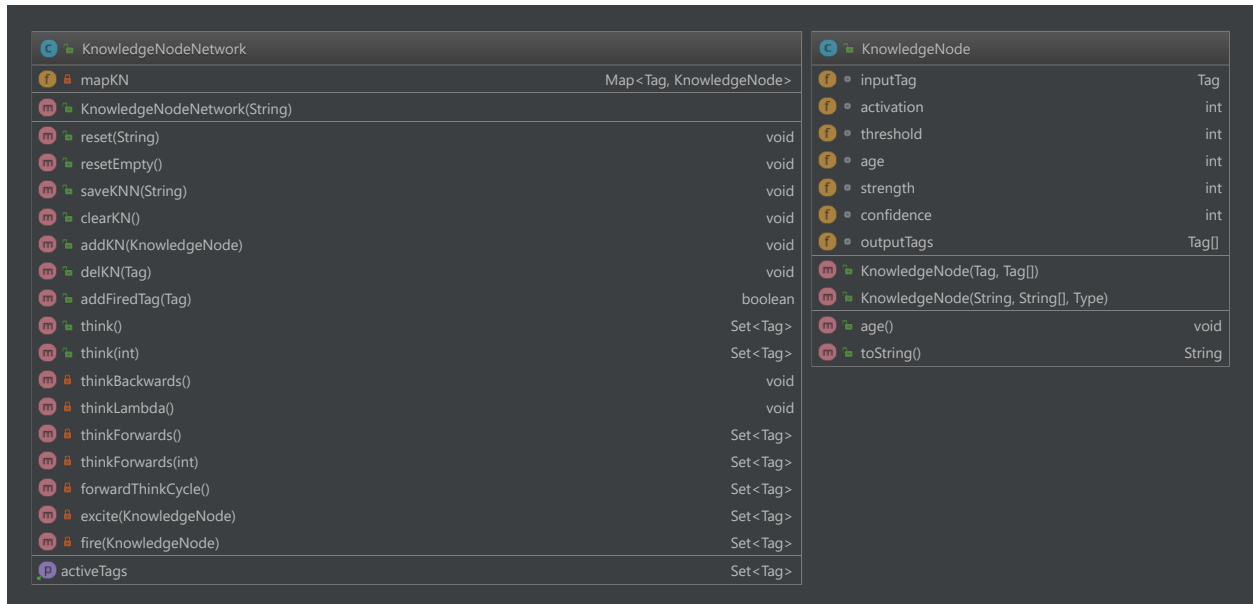


Figure A.2: UML diagram of the *knn* package.

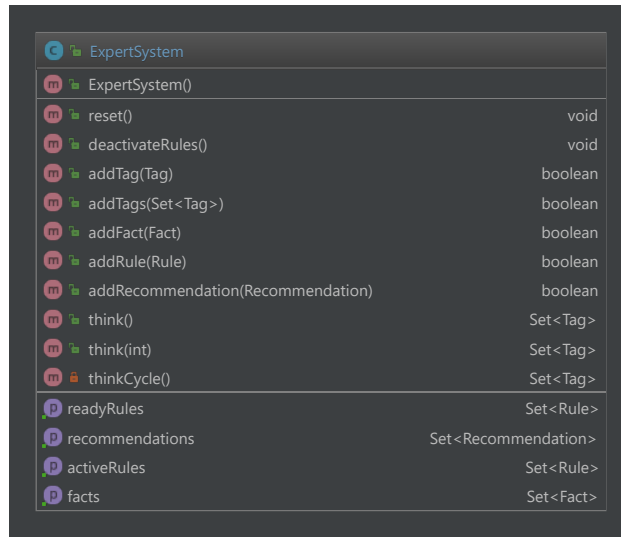


Figure A.3: UML diagram of the *es* package.

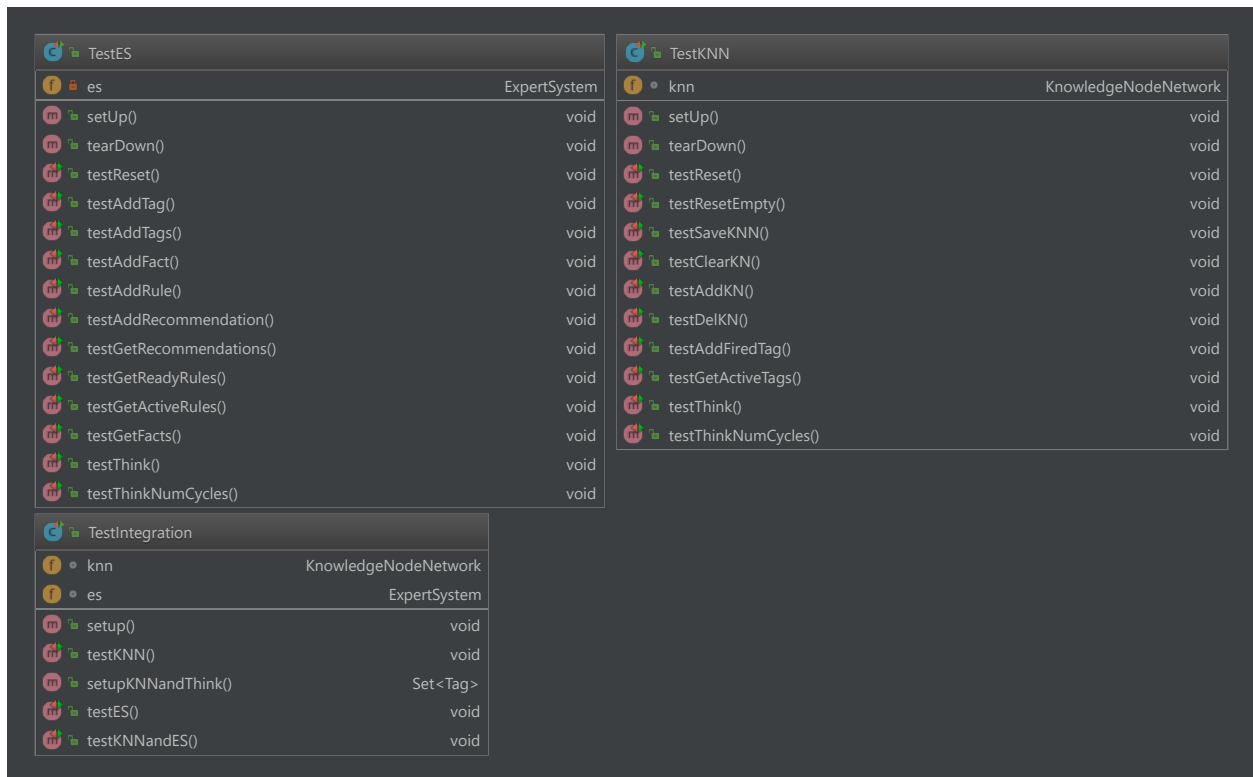


Figure A.4: UML diagram of the *test* package.

References

- [1] J. Vybihal, “Full AI model,” 2016.
- [2] C. Olah, “Conv nets: A modular perspective,” <http://colah.github.io/posts/2014-07-Conv-Nets-Modular/>, July 2014, (Accessed on 04/08/2017).
- [3] J. Vybihal, “Knowledge nodes,” 2017.
- [4] J. Palka and C. Stevens, “Neurophysiology — a primer,” 1967.
- [5] D. O. Hebb, *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.
- [6] H. L. Roediger and K. B. McDermott, “Creating false memories: Remembering words not presented in lists,” *Journal of experimental psychology: Learning, Memory, and Cognition*, vol. 21, no. 4, p. 803, 1995.
- [7] M. P. Mattson and T. Magnus, “Ageing and neuronal vulnerability,” *Nature Reviews Neuroscience*, vol. 7, no. 4, pp. 278–294, 2006.
- [8] D. C. Knill and A. Pouget, “The Bayesian brain: the role of uncertainty in neural coding and computation,” *Trends in neurosciences*, vol. 27, no. 12, pp. 712–719, 2004.
- [9] J. Vybihal and T. R. Shultz, “Search in analogical reasoning,” 1990.
- [10] D. Kahneman, *Thinking, fast and slow*. Macmillan, 2011.
- [11] J. Vybihal, “Expert systems,” 2016.
- [12] J. L. Popyack, “Object Oriented Programming,” <https://www.cs.drexel.edu/~introcs/Fa15/notes/06.1.OOP/Advantages.html?CurrentSlide=3>, 2015, (Accessed on 04/09/2017).
- [13] “Simbad 3D robot simulator,” <http://simbad.sourceforge.net/index.php>, (Accessed on 03/27/2017).
- [14] “A swarm of robots to clean up oil spills,” <https://sap.mit.edu/article/standard/swarm-robots-clean-oil-spills>, December 2010, (Accessed on 04/09/2017).
- [15] D. Gershgorn, “The US army wants to weaponize artificial intelligence — Quartz,” <https://qz.com/767648/weaponized-artificial-intelligence-us-military/>, August 2016, (Accessed on 04/09/2017).
- [16] B. Allenby, “The difficulty of defining military artificial intelligence,” http://www.slate.com/articles/technology/future_tense/2016/12/the_difficulty_of_defining_military_artificial_intelligence.html, December 2016, (Accessed on 04/09/2017).
- [17] “Open letter on autonomous weapons - future of life institute,” <https://futureoflife.org/open-letter-autonomous-weapons/>, (Accessed on 04/09/2017).
- [18] A. Kharpal, “Tech CEOs back basic income as AI job losses threaten industry backlash,” <http://www.cnbc.com/2017/02/21/technology-ceos-back-basic-income-as-ai-job-losses-threaten-industry-backlash.html>, February 2017, (Accessed on 04/08/2017).

- [19] “Bill Gates: Job-stealing robots should pay income taxes,” <http://www.cnbc.com/2017/02/17/bill-gates-job-stealing-robots-should-pay-income-taxes.html>, February 2017, (Accessed on 04/08/2017).
- [20] “Robots could demand legal rights,” <http://news.bbc.co.uk/2/hi/technology/6200005.stm>, December 2006, (Accessed on 04/09/2017).
- [21] G. Dvorsky, “A museum of robotic equipment used during the Chernobyl disaster,” <http://io9.gizmodo.com/a-museum-of-robotic-equipment-used-during-the-chernobyl-512831778>, 2013, (Accessed on 04/08/2017).
- [22] “How does mars rover Curiosity’s new AI system work? — astronomy.com,” <http://www.astronomy.com/news/2016/08/how-does-mars-rover-curiositys-new-ai-system-work>, August 2016, (Accessed on 04/07/2017).