My_Eliza,

A Multimodal Communication System

Siska Fitrianie and Dr. Drs. L.J.M. Rothkrantz
Knowledge Based Systems, Department of Mediamatics
Faculty of Information Technology and Systems
Delft University of Technology
August 2, 2002
email: s.fitrianie@kbs.twi.tudelft.nl

Abstract

My_Eliza is a computer model for a multimodal communication system, a combination of natural language processing and nonverbal communication. The development of this system is based on a famous question-answering system - QA system, Weizenbaum's Eliza [WEI66]. A human user can communicate with the developed system using typed natural language. The system will reply with text-prompts and appropriate facial-expressions. In order to communicate using a nonverbal facial display, the system should be able to process natural language and emotional reasoning. A first prototype as a proof of concept has been developed that consists of a dialog box, an emotional recognizer based on stimulus response, and a facial display generator.

To implement the dialog box, the work of Wallace, A.L.I.C.E [WAL95], has been used as a starting point. My_Eliza system has a rule engine that determines current system's affective state as reaction to the user's string input and conversation content.

1. Introduction

The way QA systems represent and retrieve information is transparent by their **memory structure**. The memory structure functions as the systems' "brain" and is the foundation of the ability level of the system to "speak" in human natural language. The QA system retrieves the information from its memory and uses syntactic and semantic analysis to output a string as an answer to the user's string input.

Eliza worked by simple pattern-matching operation and substitution of keywords. First, the system identifies the "most important" keyword occurring in the user's input string. Next, it chooses an appropriate transformation rule and its mechanism. There are two transformation rules that are associated with certain keywords. *Decomposition rule* serves to decompose a data string according to certain criteria (pattern). *Reassemble rule* serves to reassemble a decomposed string according to certain assembly specifications (reply sentence). If Eliza finds a keyword, she will pattern-match the string input against each decomposition rule for that keyword. If it matches, she randomly selects one of the reassemble rules (for that decomposition rule). Finally, Eliza uses a selected reassemble rule to construct the reply. The keyword lists, and the list of decomposition rules and reassembly rules are constructed in a script, which controls all the behavior of Eliza. Figure 1 displays an example of one unit Eliza's memory structure (asterisk sign shows that it can contain any words or phrases).

```
keyword: your
 decomposition rule: * your *
  reassemble rule: Why are you concerned about
                   my (2) ?
  reassemble rule: What about your own (2) ?
  reassemble rule: Really, my (2) ?
 decomposition rule: …
  reassemble rule: …
. . .
Example fragment:
User : What is your name?
Eliza: What about your own name?
User : Only your name, please!
Eliza: Really, my name, please?
User : Just tell me your name!
Eliza: Why are you concerned about my name?
```

Figure 1 Example of one unit Eliza's memory structure

The pattern matching operation of the original Eliza still has three major problems [SIM70]: (1) lack of anaphoric analysis, it cannot use previous question-answers to keep the continuity of the conversation content and to store information about the user, (2) lack of ability to restrict the conversation on its topic and (3) lack of ability to get the meaning beyond the sentence.

Another limitation of Eliza system is that users can only communicate with Eliza by exchanging text prompts. However beyond speech, human people can express their feelings or thoughts through the use of their body, facial expressions, and tone of voice. As indicated by

Mehrebian [KIN97], it is proved that about 55 percent of the emotional meaning of a message is communicated through the nonverbal channel, which includes gestures, postures, and facial signals. Nonverbal communication is behavior other than spoken or written communication that creates or represents meaning. Human face-to-face conversation has provided an ideal model for designing a multimodal human-computer interaction (HCI) [TAK93], [SCH00]. Characteristics of face-to-face conversation are the multiplicity and multi modality of the communication channel. Multimodal user interfaces are interfaces with multiple channels that act on multiple modalities. Conversation is supported by multiple coordinated activities of various cognitive levels. As a result communication becomes highly flexible and robust, so that failure of one channel is recovered by another channel and a message in one channel can be explained by another channel. This is the basic idea how a multimodal HCI should be developed to facilitate realistic human-machine interaction.

Nowadays, as computer acts as electronic secretaries or communication mediators, they become common entities in human society [ELL94], [NAK99]. The capability of communicating with humans using both verbal and nonverbal communication channels would be essential. This will surely make interaction between computers and humans more intimate and human-like [LEE99], [PRE01], [CAS94]. Face to face communication is inherently natural and social for human-human interaction and substantial evidence suggest this may also be true for human-computer interaction. Using human-like faces as means to communicate have been found to provide natural and compelling computer interfaces.

Eliza has shocked AI community because it gave the impression of deep semantic linguistic processing but it was in fact based on shallow language processing. Many people become emotionally involved with the QA system. Automating the recognition of users' emotion would therefore be highly beneficial in order to give a proper user reply, both in the verbal channel and in the nonverbal channel. In recent advances of QA systems, facial expression recognition and adapting life-like agents open up the possibility of automatic emotion recognition from user interaction in conversation between human and computer. Emotions are an essential part of human lives; they influence how human think and behave and how human communicate with others, and facial displays are human primary means of communicating emotion [VEL97], [SCH00]. However, there are only a few researches involving research on human emotion recognition, because it is difficult to collect a large amount of utterances that contain emotion [NAK99]. Only a few of them work in recognizing emotion from text and none of them work in facilitating emotion recognition in a QA system. Moreover, the interpretation of emotion eliciting factors is strongly situation and culture dependent [WIE99].

As a first step in achieving automatic analysis of human behavior and face-to-face communication, automated emotion recognition in human conversation between the users and a QA system has been investigated. This paper discusses the results of the research, which ensued in the development of the **my_Eliza** – an advance version of the original Eliza. My_Eliza was aimed at the design and establishment of a QA system of a ***semi*** automated emotion recognition from human user written conversation. A user or client can communicate with the system using **typed natural language**. The system will reply by text-prompts and appropriate facial-expressions.

The problem of automating emotion recognition and generating appropriate nonverbal facial displays on a QA system as defined in this research comprises into three sub-discussions: (1) automatic generation of system's reply text prompts with ability of anaphoric analysis and ability to respond the conversation based on its topic (2) semi automatic emotion recognition of user's affective state and its intensity, and (3) automatic facial display selection from a facial expressions database based on emotion analysis.

In conversation, my_Eliza displays two kinds of emotional expressions: first, related to stimulus response when she hears the utterance and second, related to cognitive processing when she realizes the situation and the conversation content to convey her reply sentences.

2. Natural Language Processing

Nowadays, a QA system is also called *chatterbot* - a short for "chatter" and "bot" [LAV96], spreading in Internet. Bot is short for "robot". A.L.I.C.E is an example of this class of programs. Tackling the three limitations of Eliza above, Wallace proposed to expand memory structure using an extended-XML (Extensible Markup Language) script specification for programming the memory structure for a QA system, called AIML (Artificial Intelligence Markup Language) [WAL95]. The most important AIML units are [BUS01]:

- **`<aiml>`**, the tag that begins and ends an AIML document.
- **`<category>`**, the tag that marks a "unit of knowledge" in the system's memory structure.
- **`<pattern>`**, the tag that contains a simple input pattern rule that matches what a user may type.
- **`<topic>`**, the tag that contains current conversation topic pattern rule.
- **`<that>`**, the tag that refers to system's previous reply as a history pattern rule.
- **`<template>`**, the tag that contains the response to a user input.

In Eliza, <pattern> tag part is namely the *decomposition rule* and <template part> tag is the reassemble rule. Figure 2 displays an example of A.L.I.C.E's memory units in a topic about name.

```
<category> <that>*</that>
<pattern>WHAT IS YOUR NAME</pattern>
<template>My <set_topic>name</set> is
         <bot name="name">.</template>
</category>

<topic name="NAME">
<category><that>MY NAME IS *</that>
<pattern>CAN I CALL YOU *</pattern>
<template><random>
 <li> <star/>? Huh! Like I've told you my name
      is <bot name="name">.</li>
 <li> You can call me whatever you like</li>
</random> </template>
</category>

<category><that>MY NAME IS *</that>
<pattern>I HATE *</pattern>
<template><random>
  <li> I don't care, you can only call me,
      <bot name="name">.</li>
  <li> Why? A sad memory perhaps?</li>
</random> </template>
</category>
</topic>
…
Example fragment:
User : What is your name?
Alice: My name is Alice
User : Can I call you Madonna?
Alice: Madonna? Huh! Like I've told you my name
       is Alice.
User : I hate that name.
Alice: Why? A sad memory perhaps?
```

Figure 2 An example of A.L.I.C.E's memory units

<set> and <get> tags are used to store information during conversation. See also [BUS01], and [WAL95]. A.L.I.C.E has much more possibilities of reply sentences based on their topic and history. Using AIML gives the possibility to create new content by a dialog analysis.

The matching operation is word-by-word, not category-by-category. The algorithm searches the best match pattern by ensuring that the most specific pattern matches first basically it finds the longest pattern matching an input. If there are two identical patterns but the later contains the same <that> tag, then it will take precedence over the other categories, if inside <that> tag matches the previous response. Any categories that are contained within a <topic> tag will be searched first if the current topic matches it. If neither of above is true, there is a default category with <pattern>*</pattern>. We used this AIML schema to build my_Eliza's memory structure.

3. Nonverbal Communication

This section deals with emotion reasoning and facial display generator. The main goal here is to explore the issues of design and implementation of a nonverbal QA system that could recognize the user's emotion and show a proper facial display accordingly. In general, three steps can be distinguished in tackling this issue: (1) define which and how many emotions can be recognized by the system, (2) define mechanisms for extracting emotion-eliciting factors in the observed text prompt, which devise the categorization mechanism and the emotion interpretation mechanism, and (3) define some set of categories of emotions that we want to use for facial displays classification and facial displays generation mechanism.

Currently, the interpretation of the emotion-eliciting factor is still semi automatic since we assume to use the memory structure approach of Weizenbaum's or Wallace's pattern matching operation. The memory structure of this approach does not store the semantic meaning of the text. It needs human intervention to interpret the affective semantic meaning.

4. Emotion Classification

How many and what kind of emotional expressions are to be treated in a QA system are interesting but difficult issues. In this research we investigate three classification methods:

1. Reddy's [RED01] basic emotions: every emotion is either pleasant or unpleasant and every emotion has a varying intensity regarded as either shaping one's goals or reflecting one's goals.
2. Ekman and Friesen's [EKM75] seven universal emotions: neutrality, happiness, sadness, anger, fear, disgust, and surprise, in terms of facial expressions and mainly concentrated on primary or archetypal emotions, which are universally associated to distinct expressions.
3. Ortony, Clore and Collins theory's twenty-four emotions [BAZ01], [ELL93] (OCC's theory 1988, see table 1). It is based on grouping human emotions by their eliciting conditions events, their consequences of their action, and their selections of computational implementation. They are resulted in three branches: (1) *Attraction* relates to emotions that are arising from aspects of the object, (2) *Consequences of event* relates to reaction of others' fortunes and (3) *Attribution* relates to approval of self or other. In addition, there is a *compound* class that involves the emotions of gratification, remorse, gratitude and anger.

Table 1 Twenty-four emotion types according to OCC's theory

Name and Emotion Type
Joy: pleased about an event **Distress**: displeased about an event
Happy-for: pleased about an event desirable for another **Gloating**: pleased about an event undesirable for another **Resentment**: displeased about an event desirable for another **Sorry-for**: displeased about an event undesirable for another
Hope: pleased about a prospective desirable event **Fear**: displeased about a prospective undesirable event
Satisfaction: pleased about a confirmed

Name and Emotion Type
desirable event **Relief**: pleased about a disconfirmed undesirable event **Fears-confirmed**; displeased about a confirmed undesirable event **Disappointment**: displeased about a disconfirmed desirable event
Pride: approving of one's own act **Admiration**: approving of another's act **Shame**: disapproving of one's own act **Reproach**: disapproving of another's act
Liking: finding an object appealing **Disliking**: finding an object unappealing
Gratitude: admiration + joy **Anger**: reproach + distress **Gratification**: pride + joy **Remorse**: shame + distress
Love: admiration + liking **Hate**: reproach + disliking

Since classifications of some emotion eliciting factors are in a gray area, in this research, we add one emotion type: uncertainty.

4.1. Emotion Eliciting Factor Extraction

Most of developed systems that are able to devise emotion-eliciting factor information still need manual human intervention. Following three experiments dealing with representing and extracting emotions' information on the system's memory structure and how we map them in my_Eliza:

1. Emotive lexicon dictionary look-up parser.
This approach uses a list of lexicons associated to different type of emotions. Those lexicons, which are composed by words or phrases, are selected from the way human people expresses their feelings with its intensity. The system uses a shallow word matching parser to extract affective state from the context. Elliott [ELL95], [ELL92], [ELL93] used this approach for his model of a multi-agent world where each agent is able to reason about emotion episodes that take place in one another's lives. He used an extended base lexicon of spoken phrases that includes 198 emotion words associated with twenty-four OCC's theory emotion types. Those words describe relationship, mood and emotional intensity. Each emotion type has a set of eliciting conditions. When the eliciting conditions are met, and various thresholds have been crossed, corresponding emotions result. The system applies minimal the detection of user's emotional inflection. Using this approach, it allows the user to teach the computer keywords in a new vocabulary relatively quickly and the system remains understandable no matter in which context the user is.

My_Eliza uses this approach to extract emotion-eliciting factor information in the text prompt in the conversation both of the user's string input and the system's reply sentence. Since the first prototype is dedicated as a "proof of concept", only six universal emotion types (Ekman's) will be used in emotive lexicons classification instead of twenty-four OCC's theory emotion types. We define six dictionaries containing lexicons in the following form: `[<lexicon>: <intensity value>]` with `<intensity value>` is an integer value [1..3].

We also define six affective counters C for each emotion type. The parser parses the sentence word-per-word against the dictionary. If it finds the same emotive lexicon in the dictionary, it will calculate the counters using following equations:

$$\forall \text{ Lexicon } l_i \in d_i \mid C_{i(t)} = C_{i(t-1)} + I_i \cdot s \;;$$

i = active emotion type

I = intensity level; s = summation factor

$$\forall\, j \neq i \mid C_{j(t)} = C_{j(t-1)} - \text{distance}[j, i]$$

j = {happiness, sadness, anger, fear, disgust, surprise}

For the first prototype we use Hendrix and Ruttkay's distance values between expression emotions ([HEN98], table 2) for `distance[j, i]`. The result of this calculation is the candidate of affective state both for the user and the system, which is taken from the emotion type with the highest level of all counters.

Table 2 Distance value between emotions [HEN98]

	Happiness	Surprise	Anger	Disgust	Sadness
Happiness	0	3.195	2.637	1.926	2.554
Surprise		0	3.436	2.298	2.084
Anger			0	1.506	1.645
Disgust				0	1.040
Sadness					0

2. Emotive labeled memory structure extraction.
This approach labels each unit of memory structure with one or more of the emotions types. Most of the examples for systems using this approach are automatic story telling systems and automatic digitizer for cartoon movies. Each dialog sentence of each actor is labeled with an emotion type and decomposed in its phonological representation. Therefore, the system can show appropriate intonation and nonverbal display when it reads the dialog. Pelachaud et.al. [PEL94] used it in their research by assuming the input as a file containing an utterance already decomposed and written in its phonological representation with its prosody in its bracketed elements. At each input, it specifies the desired affectual parameters and their intensity. Using this approach means to sidestep the issue of emotion recognition. The modeling of affect is not meaning based and it needs human manual work to label each memory unit.

We label my_Eliza's memory structure by adding two additional tags in AIML schema: <affect> tag that labels the user's affective situation and <concern> tag that labels the system's reaction situation. Inside those two new tags, based on Reddy's basic emotion we define four-possibility emotive situation type: (1) positive "+", (2) negative"-", (3) joking "#", and (4) normal/any "*". Figure 3 (below) displays an example of my_Eliza's memory structure units.

```
<category> <affect name=”*”>
<pattern>WHAT IS YOUR NAME</pattern>
<that>*</that>
<template><think><setconcern>+</setconcern>
  <setaffect>+</setaffect></think>My
  <set_topic>name</set> is <bot name=”name”>.
</template>
</affect>
</category>

<topic name="NAME">
<category> <affect name=”*”>
<that>MY NAME IS *</that>
<pattern>YOUR *</pattern>
<template><random>
 <li><think><setconcern>#</setconcern>
  </think>Your <star/> also!</li>
 <li><think><setconcern>+</setconcern>
  </think>Mind that, what is your name?</li>
</random> </template>
</affect></category>

<category> <affect name=”*”>
<that>MY NAME IS *</that>
<pattern>I HATE YOUR *</pattern>
<template><random>
 <li>><think><setconcern>-</setconcern>
  </think>Why? A bad memory perhaps?</li>
 <li>><think><setconcern>-</setconcern>
  </think>Why?</li>
</random>
<think><setaffect>-</setaffect></think>
</template>
</affect></category>

<category> <affect name=”-”>
<that>MY NAME IS *</that>
<pattern>YOUR *</pattern>
<template>
 <li><think><setconcern>+</setconcern>
  </think>Mind that, what is yours?</li>
 <li><think><setconcern>-</setconcern>
  </think>Don’t be a witty!</li>
</random> </template>
</affect></category>
</topic>
…
```

Figure 3 Example of my_Eliza's memory structure units

Using the memory structure units in figure 3, we expect my_Eliza is able to converse with human user in positive situation (table 3) and in negative situation (table 4).

Table 3 Example of fragment in positive situation

User:	What is your name?	 My_Eliza
 My_Eliza:	Your name is lovely.	
User:	My name is Eliza.	 My_Eliza
 My_Eliza:	Your name is lovely also!	

Table 4 Example of fragment in negative situation:

User:	What is your name?	 My_Eliza
 My_Eliza:	My name is Eliza.	
User:	I hate your name.	 My_Eliza
 My_Eliza:	Why? Bad memory perhaps?	
User:	Your name reminds me to my enemy.	 My_Eliza
 My_Eliza:	Mind that, what is yours?	

Using extra two tags, my_Eliza has the possibility to have a bigger memory structure than A.L.I.C.E. In matching-operation algorithm, the memory structure with the same situation type will be checked first. If none of them is matches, the system picks a category with a normal situation type (*).

3. Goal based emotion reasoning.

This approach sets some goals, principles, preferences and moods in the system. Extraction emotion-eliciting factors leading to emotions falls into four major categories: those rooted in the effect of an event on the goal of the system, those rooted in the standard and principle invoked by an act of the user, those rooted in tastes and preferences with respect to an object (including the user treated as an object), and lastly a selected combination of the first three categories. Another way to view these categories is that they are rooted in the system's assessment of the *desirability or undesirability* of some event, the *praiseworthiness or blameworthiness* of some act, the *attractiveness or unattractiveness* of some object, or selected combinations of these assessment. Elliot [ELL93a] also used this approach for his model to simulate social interactions between agents in incorporated models of individual affect and personality. Each agent interprets situations that are characterized in terms of the way they may or may not meet the eliciting conditions of emotions. Agents use a case based heuristic classification system to reason about the emotions of other agents' personalities that will help them to predict and explain future emotion episodes involving observed agents. Embodied in the simulation system, Elliott used a set of rules for the mapping from four categories emotion-eliciting factors above into twenty-four OCCs theory emotion types.

Mapping to my_Eliza, we define the system's goals, affective status and preferences (GSP) while she converses with the user. Several goals of my_Eliza are:

- *Answering questions* – if the user asks something, my_Eliza's goal is to answer it.
- Persuasive agreement – if the user persuades to do something or invites my_Eliza to do something, my_Eliza's goal is to show whether she agrees or not.
- *Topical focus* – to keep on conversing on the same topic and beware if it is changing.
- *Explanation statements* – to reply the user's statements that require specification and explanation.
- *Reflecting feeling* - to keep consistent with the user's current affective state.
- *Alignment* - to keep consistent with the system's current reaction affective state (system's status) and system's preferences.

The system recognizes the dialog using a dialog scheme adopted from [CAR95]. By distinguishing the by distinguishing a dialog state as a certain dialog act like a question, statement, acknowledgement, or pause, the system has to know which goal to pursue. Whether a certain goal is appealing or not appealing may influence the system's affective state. We also define my_Eliza's preference as the personal data about the system and can be used during conversation, for example: her name, birthday, the things she likes or hates, and so on. To be fair, using <set> tag the system also stores the user's personal data during conversation, for example the user's name, birthday, favorite stuff, personal data about family and so on. These data about system's GSP and user's personal data can be used for pragmatic analysis when the system constructs the reply sentence and defines its current affective status.

4.2. Emotion Recognition

For the activation of an emotion, [ELL93], [VEL98], [PRE01], and [BAZ01] proposed the use of threshold values by counting all associated elicitation factors, the excitatory (positive) and inhibitory (negative), from other emotions. They used an activation level range [0, max] where *max* is an integer value determined empirically. All emotions are always active, but their intensity must exceed a threshold level before they are expressed externally. The activation process is controlled by a knowledge-based system that synthesizes and generates cognitive-related emotions in the system.

We define six affective thermometers classified by six Ekman's universal emotion types. These thermometers observe the affective state of the system as reaction to the user's string input and the dialog content – the system's reaction affective state. If an emotion is active, the system calculates all of thermometers *T*, with the following equations:

$T_{i(t)} = T_{i(t-1)} + \boldsymbol{I_i} \cdot \boldsymbol{s}$

i = an active emotion type

s = summation factor; I = intensity

$\forall\ j \neq i \mid T_{j(t)} = T_{j(t-1)}$ - distance[j, i]

Where j = {happiness, sadness, anger, fear, disgust, surprise}

We also use Hendrix and Ruttkay's distance values to calculate those equations. The system takes the highest degree of all thermometers as the most dominant emotion.

To determine the system's affective state we formulate two knowledge based systems: (1) determines the system's reaction affective state as stimulus response to the user's input string and (2) determines the system's reaction affective state as the result of cognitive process of the conversation content to convey its reply sentence. We have defined a set of rules that specify the emotion recognition process of the system. We call these rule-sets ***preference rules***, since they indicate preferences to exhibit system's "preference" reaction affective state rather than performing explicit actions, such as facial displays. Every rule in the set defines conditions of emotion eliciting factors and the affective thermometers to activate the rule and a preference that is expressed upon activation. The result from each knowledge-based system is one of twenty-four OCC's theory emotion types with addition of two emotion types: normal and uncertainty, for example:

1. Preference rule for stimulus response:

This rule will fire the preference first reaction **joy** if the following conditions are met:

```
→ user is happy,
→ user asks question,
→ situation type of user is not negative,
→ current maximum system's affective
  thermo is happy.
```

In this case my_Eliza will answer any questions from the user joyfully, because she enjoys the situation and she met the goal: making the user feel happy.

2. Preference rule for cognitive process:

This rule will fire the cognitive processed preference **resentment** if the following conditions are met:

```
→ user is sad,
→ system's reply is sad,
→ situation type of user is joking,
→ situation type of the system is negative,
→ current maximum system's affective
  thermo is sad.
```

Here my_Eliza does not like the user makes a joke while she feels sad.

4.3. Facial Display Generator

In most of the works in facial display generation are used one to one corresponding facial display and emotions, distinguished by intensity [ELL93], [PRE01]. The other works used the correspondence between communication categorization and Ekman & Friesen's Facial Action Coding System (FACS) [TAK93] and between emotions with FACS [PEL94]. FACS is a notation to describe visible facial expression based on anatomical studies; how a feature is affected by specifying its new location and the intensity of changes.

For the first prototype we use one to one corresponding facial display and emotion. We use twenty-two smiley nonverbal facial display classified by eight emotion types (neutrality, happiness, sadness, anger, surprise, fear, disgust and uncertainty) and three level of intensity (LOW, MEDIUM, HIGH) except for neutrality. Since the system's reaction affective state may be one of twenty-four OCC's theory emotion types we cluster every those emotion types into six Ekman's universal emotion types. We cluster normal into neutrality.

5. Design

The architecture of my_Eliza is illustrated in figure 4, which takes the idea of message passing on a blackboard system. The message flow and message process are always on the blackboard. If a new message comes, it will be analyzed, synthesized, and the result will always be put back on the blackboard. In my_Eliza, the message is the user's **string input** and the results are the **reply sentences and facial displays**. My_Eliza works by the following steps:

1. Generating a stimulus-response nonverbal signal

- User types *a string input* and puts it on the blackboard system.
- The *Parser* parses the input into words and puts it on the list on the blackboard system.
- The *Lexical Analysis layer* normalizes the string input by eliminating incorrect or incomplete words or phrases and checking relations between words or phrases. This layer puts the result on the blackboard system.
- The *Affective State Analysis layer* activates its two sub layers: *Emotive Lexicon Dictionary Parser* and *Emotive Labeled Memory Structure Extraction*. The *Emotive Lexicon Dictionary Parser layer* identifies the emotive lexicons from the user's input and the reply sentence (after the system has constructed the reply sentence). The *Emotive Labeled Memory Structure Extraction layer* extracts the label from the system's memory unit. Those results are put on the blackboard system.
- The *Syntactic-Semantic Analysis layer* performs a pattern matching operation based on the user's string input pattern to add the user's affective information on the blackboard system. In this step, the system starts to work in parallel with the process to construct system's reply sentence. As a result, *Syntactic-Semantic Analysis layer* performs pattern-matching operation to generate system's reply sentence and puts its candidate on the blackboard system.
- The *Affective Attributing Analysis layer* activates its sub layer, the *Concern of the other Analysis layer*, to perform emotion-based reasoning to deliver current system's reaction affective state and put it on the blackboard system. This analysis based on stimulus response and directly related to the user's input string.
- The *Intensity Analysis layer* processes the message and calculates the current system's affective 'thermometers' and puts the calculation result on the blackboard system.
- The *Facial Selection* selects my_Eliza's facial display and puts the selection on the blackboard.
- The *Wrap Process* delivers and displays my_Eliza's **stimulus-response facial display** or the first facial display.

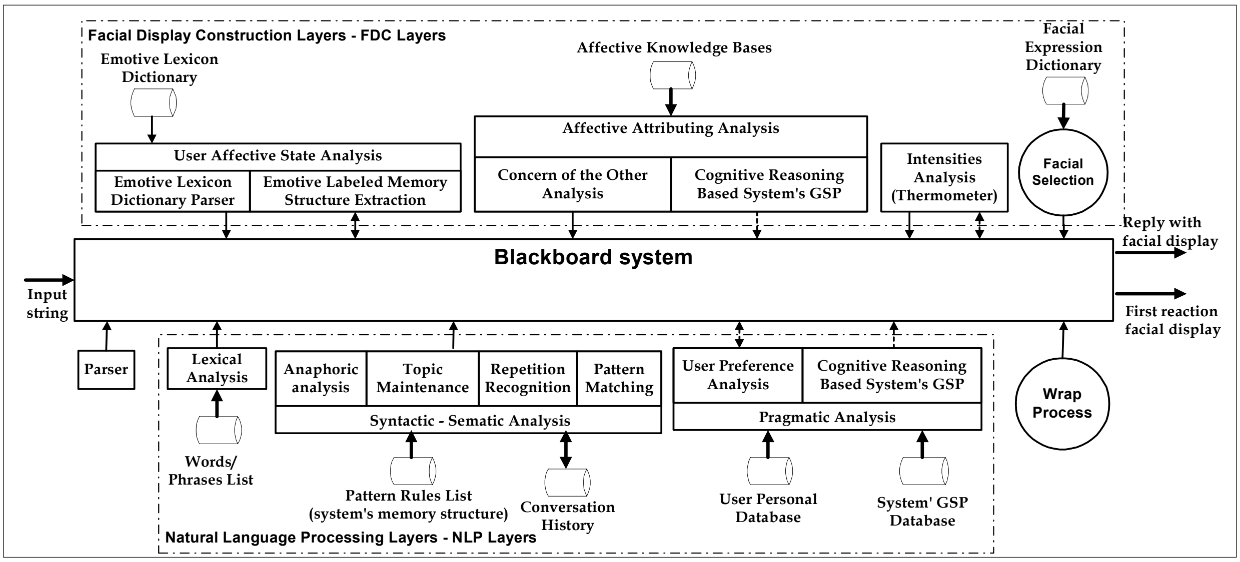

Figure 4 My_Eliza's blackboard system

2. Constructing the candidate of reply sentence and generating a cognitive-processed facial display:

- The *Pragmatic Analysis layer* processes the candidate reply sentence and puts the result on the blackboard system. This layer reviews the candidate of reply sentence whether it violates user's preference and/or system's goals, status and preferences – system's GSP. If it does, the result will be sent back to the *Syntactic-Semantic Analysis Layer* to get new a reply sentence.
- The *Affective Attributing Analysis layer* activates its sub layer, the *Cognitive Reasoning layer*, to perform emotion-based reasoning to deliver current system's reaction affective state and put it on the blackboard system. This analysis is based on cognitive processing of system's GSP, system's reply sentence and user's affective state.
- Again, the *Intensity Analysis layer* processes the message and calculates the current system's affective 'thermometer' level. This layer also puts the result of calculation on the blackboard system.
- The *Facial Selection* selects my_Eliza's facial display and puts the selection on the blackboard.
- The *Wrap Process* delivers and displays my_Eliza's **reply sentence** and **cognitive-processing result facial display**.

6. Implementation

There are three incremental implementation layers: (1) create a dialog box that can engage in human conversation based on typed natural language and recognize the user's affective state and the system's reaction affective state, (2) build a stimulus-response of facial displays based on spontaneously spinal brain reasoning on user's string input, (3) build a cognitive processor of facial displays based on anaphora analysis, pragmatic analysis, dialog content and system's goals, status, and preferences.

Currently, we are in the second implementation layer and the result is called my_Eliza prototype-1. We use Program D A.L.I.C.E [WAL95] as a starting point to build my_Eliza's dialog box. Program D A.L.I.C.E is written on Java Development Kit version 1.3 and XML, therefore we use a compiler and classes contained in the same languages for my_Eliza prototype-1. My_Eliza's dialog box contains many packages most of them derive from Program D A.L.I.C.E. This program has provided a robust client server and multi user communication. Therefore, My_Eliza is also controlled by a collection of autonomous client-server communication via TCP/IP. The user can communicate with my_Eliza's server through the HTTP server. My_Eliza's server provides the blackboard system. My_Eliza uses the **Java expert system shell** (Jess) for *affective attributing* knowledge based system shell [FRI00]. Jess is a rule engine and scripting environment written entirely in Java. The script of the knowledge base is written in .clp.

Currently, my_Eliza prototype-1's emotive lexicon dictionary contains: 48 lexicons for happiness, 170 lexicons for sadness, 34 lexicons for surprise, 33 lexicons for fear, 93 lexicons for disgust, and 69 lexicons for anger. This prototype has 1953 categories in its list of pattern rules. Its affective knowledge bases contain 77 preference rules of stimulus response knowledge base and 151 preference rules of stimulus response knowledge base. We can add these databases and knowledge bases easily even while the server is still running. With this benefit we can build the system's knowledge base incrementally by trial error. Figure 5 displays the main page of my_Eliza prototype-1 when she felt sorry-for the user's (namely Siska) misfortune.

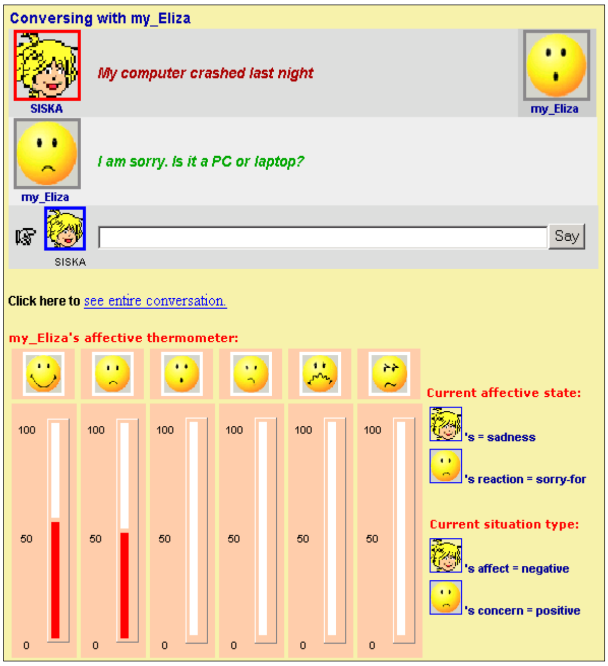

Figure 5 My_Eliza prototype-1 main page

7. Conclusion

Surveying the literature about the emotion recognition process leads to the conclusion that none of the research is particularly in the work of recognizing emotion in a QA system. However, from the work of researchers in the fields of multi agents system, emotion recognition from human speech intonation, automated character animation system and communicative facial display system, could give inspiration which allows us to fit in our approach with the fields. However from most of the work still need human manual intervention. A pragmatic advantage of using AIML form and preference rules to exhibit system's behavior is that the systems' memory structure and its behavior can be extended easily. One of future works in this research is to extend the rule-sets in system's affective knowledge bases. Current rule-sets that we have implemented in the system do not know the correlation emotion of current dialog and the emotion of entire conversation content. The rules-sets could also be made temporal.

We can make specific rules for the opening of the conversation, during discussion or the end of the discussion. By adding more emotion eliciting-factors in the rule-sets, extra rules are needed for new eliciting factors.

Additionally the server interface should have extended functionality to show the thinking process of the system. Realistic virtual environments not only include believable appearance and simulation of the virtual world but also imply the natural representation of participants. That can be fulfilled by visualization of human character embodiment with animation. Moreover, using more possible facial displays, the system is able to convey many different kinds of emotion as different social situation arise. It needs to explore several ways so that real-time, animated, and virtual human characters can be given more intelligence and communication skills, therefore they can act, react, make decisions, and take initiatives. In a similar fashion, the system should be able to communicate with a broad range of conversation topics and it should be able to visually support these conversations with an equally broad range of emotion and expressions behaviors. The system also should have the ability to learn from conversation history. These additions to the system will be valuable assets to add new memory structure units and to add rule-sets of the system to generate its affective knowledge bases autonomous.

References

[BUS01] Bush, Noel, *AIML Specification ver. 1.0.1*, http://Alicebot.org/TR/2001/WD-aiml/, 2000.

[BAZ01] Bazzan & Bordini, A Framework for the Simulation of Agents with Emotions, Report on Experiments with the Iterated Prisoner's Dilemma, AGENT'01, Communication of ACM, p292-p299, Canada, 2001.

[CAR95] Carletta, J., et. al., *The Coding of Dialogue Structure in a Corpus, in Corpus-based Approaches to Dialogue Modeling,* 9th Twente Workshop on Language Technology, p25-p34, University of Twenty, 1995.

[CAS94] Cassell, J. et.al. *Animated Conversation: Rule Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversation Agents*, in SIGGRAPH'94, 1994.

[EKM75] Ekman & Friesen, *Unmasking the Face*, Prentice Hall, New Jersey, USA, 1975.

[ELL92] Elliott, Clark, *Thesis Report*, *The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System*, Ph.D. Desertation, Northwestern University, The Institute for the Learning Sciences, Tehnical Report No.32, 1992.

[ELL93] Elliott & Siegle, *Variables Influencing the Intensity of Simulated Affective States*, In AAAI Technical Report for Spring Symposium on Reasoning about Mental States: Formal Theories and Applications, 58-67, American Association for Artificial Intelligence, 1993.

[ELL93a] Elliott, *Using the Affective Reasoner to Support Social Simulations*. In Proceeding of the 13th International Joint Conference on Artificial Intelligence, 194-200, Chambery, 1993.

[ELL95] Elliott, *Draft: Components of two-way emotion communication between humans and computers using a broad, rudimentary, model of affect and personality*, http://condor.depaul.edu/~elliott/ar.html, 1995.

[FRI00] Friedman-Hill, Ernest, *Jess: Rule Engine of the Java Platform*, Sandia National Laboratory, http://herzberg.ca.sandia.gov/jess/, USA, 2000.

[HEN98] Hendix & Ruttkay, *Exploring the Space of Emotional Faces of Subjects without Acting Experience*, ACM Computing Classification System: H.5.2, I.5.3, J.4, ftp://ftp.cwi.nl/ pub/CWIreports/INS/INS-R0013.ps.Z, 1998.

[KIN97] King, Donnell, A., *Nonverbal Communication*, http://www2.pstcc.cc.tn.us/~dking, 1997.

[LAV96] Laven, Simon, *The Simon Laven Page, Chatterbot Central*, www.simonlaven.com, 1996.

[LEE99] Lee, Kwanyong, *Integration of Various Emotion Eliciting Factors for Life-Like Agents*, ACM Multimedia'99, Part 2, p155-p158, 1999.

[NAK99] Nakatsu, Ryohei, et al. *Emotion Recognition and Its Application to Computer Agent with Spontaneous Interactive Capabilities*, Creativity and Cognition 99, Communication of ACM, p135-p143, UK, 1999.

[PEL94] Pelachaud et.al., *Generating Facial Expression for Speech*, Dept. of Computer and Information Science, University of Pennsylvania, 1994.

[PRE01] Prendinger & Ishizuka, *Social Role Awareness in Animated Agents*, AGENTS'01 ACM, p270-p276, 2001.

[RED01] Reddy, W.M., *The Navigation of Feeling, A Framework for the History of Emotion*, Cambridge University Press, Edinburgh, 2001.

[SCH00] Schiano et.al., *Face to Interface: Facial Affect in (Hu)Man and Machine*, CHI Letters, Communications of the ACM, Vol. II No. 1, p193-p200, 2000.

[SIM70] Simmons, R.F., *Computational Linguistic, Natural Language Question Answering System: 1969*, Communications of the ACM, Vol. XIII No. 1, p15-p29, January 1970.

[TAK93] Takeuchi & Nagao, *Speech Dialogue with Facial Displays: Multimodal Human-Computer Conversation,* Agent Interfaces, Section VII, p572-p579, 1993.

[VEL98] Velasquez, J.D., *Modelling Emotions and Other Motivations in Synthetic Agents*, American Association for Artificial Intelligence (AAAI), www.aaai.org, 1997.

[WAL95] Wallace, Richard, *Alicebot*, http://www.Alicebot.org, 1995.

[WEI66] Weizenbaum, J., *ELIZA – A Computer Program for the Study of Natural Language Communication between Man and Machine*, Communication of the ACM 9(1): p36-p45, 1966.

[WIE99] Wierzbicka, Anna, *Emotional Universal*, Language Design 2, p23-p99, Australia, 1999.