Designing a Chat-bot that Simulates an Historical Figure

Emanuela Haller, Traian Rebedea
Faculty of Automatic Control and Computers
University Politehnica of Bucharest
Bucharest, Romania
emanuela.haller@cti.pub.ro, traian.rebedea@cs.pub.ro

Abstract—**There are many applications that are incorporating a human appearance and intending to simulate human dialog, but in most of the cases the knowledge of the conversational bot is stored in a database created by a human experts. However, very few researches have investigated the idea of creating a chat-bot with an artificial character and personality starting from web pages or plain text about a certain person. This paper describes an approach to the idea of identifying the most important facts in texts describing the life (including the personality) of an historical figure for building a conversational agent that could be used in middle-school CSCL scenarios.**

Keywords— Chat-bot, Conversational Agent, Information Extraction, Parsing, Wikipedia

I. INTRODUCTION

A chat-bot is a computer program designed to simulate an intelligent conversation with a human user via auditory or textual methods. Most of the chat-bots are designed for engaging in small talk and their personalities are created by the programmer. A good introduction in designing chat-bots using the current state of the art, which is mainly using rules written in AIML (Artificial Intelligence Markup Language) [1] or ChatScript [2], is provided by [3]. Some examples of good chat-bot programs are: Cleverbot [4], Do-Much-More [5], Suzette [6], and Rosette [7].

The Loebner Prize is an annual competition in Artificial Intelligence that awards prizes to the chat-bot considered by the judges to be the most human-like. The format of the competition is that of a standard Turing Test. This is the first formal instantiation of a Turing Test. In 1950, Alan Turing asked the question “Can a Machine Think?” [8]. He answered in the affirmative, but a second question arose: “If a computer could think, how could we tell?”. Turing’s suggestion was that if the responses from the computer were indistinguishable from that of a human, the computer could be said to be thinking.

An even more difficult challenge begins when trying to design an artificial entity which should not only be human-like, but should also act as a specific human being. Also, the program has to be able to learn from its prior experiences, in order to make the conversation even more realistic.

This paper presents a method for building a chat-bot that can simulate an historical figure. The purpose is to provide a generic solution to this problem, so the goal is not to simulate the life and behavior of particular person. Rather, we would like the bot to automatically acquire as much information about the life and personality of the simulated person so that it can act accordingly. Thus, the intention is to develop an application that can receive as “input” a plain text or a web page about the historical figure and has as “output” a trained conversational agent which is able to answer all kind of questions about the life experience of that specific person.

The human personality is to some degree inborn, but also to some degree it is molded by the individual’s life and social interactions. In consequence, this artificial entity can be perceived as a human with no inborn personality, but only with a personality molded by the facts the program has been able to extract from the person’s life. This type of chat-bots can be successfully used in learning situations, wither formal or informal, for example as moderators in history classes where students use CSCL chats to discuss about each historical character, but also in museums as a means to offer personalized information to young people by engaging them in conversations.

The next part of the paper is related to how a program can automatically extract structured data from plain text in order to “import” the life events and thus the personality of the artificial entity. Then we proceed to discuss about the actual implementation of the conversational agent starting from this knowledge base for each individual person. At the end, some results are highlighted and commented using a specific historical figure.

II. EXTRACTING STRUCTURED DATA

The first concern is to identify open and, to some extent, reliable sources of information to extract your knowledge from. Some good choices are Wikipedia (http://www.wikipedia.org/) and DBpedia (http://dbpedia.org/About), as they provide correct information for a wide range of historical figures. Thus, the application designed to build the knowledge base for each person first asks the user to provide the two links: one to the Wikipedia page of the character and one to the DBpedia resource. This is the only information required from the user in order to be able to simulate the life and personality of the desired character.

As a generic representation of knowledge, each data will be recorded as a fact, which is a triplet:

(subject, action/verb, object)

The fields subject, action/verb, and object should contain no whitespaces, unless the entire value is encased in double quotes. Thus embedded blanks in the analyzed text are replaced with underscores ('_'). When defining facts, the verb may be omitted if the relation has the meaning of *has-a* or *is-a*. For example, the triplet (surname, is, Hitler) means that the surname of the character is Hitler. Moreover, the same subject can refer to different objects. In this case, if we want to find out the wars in which the character was involved, we did not create the following two facts:

(World_War_I, is-a, war)
(World_War_II, is-a, war)

Rather, we have created three different facts:

(wars, number-of, 2)
(war-1, has-name, World_War_I)
(war-2, has-name, World_War_II)

This way, it is very simple to know how many wars the character has taken part into and we can quickly access information about each one of them. So, the generic form for adding facts to the knowledge base is:

(subjects, number-of, n)
(subject-i, verb/action, object) , i=1..n

From now on we will use the person of Adolf Hitler for exemplification, without losing the generality of the approach.

For extracting facts from text, we have used the Stanford CoreNLP libraries [9], which provide a set of natural language analysis tools which can take raw English language text input and perform lemmatization, POS tagging, markup the structure of sentences in terms of phrases and word dependencies, and many other facilities.

The most useful functionality was the Stanford typed dependencies representation, which was designed to provide a simple description of the grammatical relationships in a sentence. The dependencies map straightforwardly onto a directed graph representation in which the words in the sentence are nodes and grammatical relations are edge labels [10]. Another useful tool is the part-of-speech (POS) tagger, which assigns parts of speech to each word in the analyzed text which offers a Java implementation of the log-linear POS taggers [11].

The last linguistic resource used in this stage is WordNet (http://wordnet.princeton.edu/), a large lexical database for English. Nouns, verbs, adjectives and adverbs are grouped into sets of synonyms, each expressing a distinct concept. So, it is not only useful to find out the meaning of a word, but also to find synonyms for it.

Next, we shall proceed to discuss how these tools are used for extracting structured data that represents knowledge from Wikipedia and DBpedia.

A. *Using Information from DBpedia*

DBpedia is a community effort to extract structured data from Wikipedia and to make this information available on the Web. It has the disadvantage that for each personality there is only little useful information already extracted, which does not suffice for reaching our target.

Different types of information were imported from DBpedia, some directly from the resource associated to the character, and some from other resources which are linked through various relations to the main page of the figure under consideration.

For example, if one finds out from the character's resource that the he was married with Eva Braun, one will want to parse this resource as well, in order to find out more information about her. In consequence, whenever you find a reference to another resource about a relevant subject to the character, this is also downloaded analyzed to extract useful information from it and enrich the knowledge base.

From DBpedia, you can access raw data in different formats, including CSV (comma-separated values), RDF (resource description framework), OData (open data protocol) and Microdata. In this application, RDF files are used which are exported using the N3/N-Triples data format. In order to obtain the N3/NTriples file, if the main link is http://dbpedia.org/page/Adolf_Hitler, you need to download the resource from the following link http://dbpedia.org/data/Adolf_Hitler.rdf.

The representation of data makes it very easy to parse. For example, from following RDF line can be very easily transformed to our triple representation format that uses natural language rather than tags to represent that the first name of the character is "Adolf":

<foaf:givenName xml:lang="en">Adolf</foaf:givenName>

As previously mentioned, for interesting subjects connected to out character that have their own pages on DBpedia, we are interested to find out the links to those pages. This can be achieved by identifying the links under the "rdf:resource" attribute identify other DBpedia resources:

<dbpprop:spouse
df:resource="http://dbpedia.org/resource/Eva_Braun" />

For each historical figure, several basic information was imported from DBpedia pages, like name, surname, a briefly introduction about the character, some data about military career, political orientation, birth place and date, death place and date, etc.

B. *Extracting Information from Wikipedia*

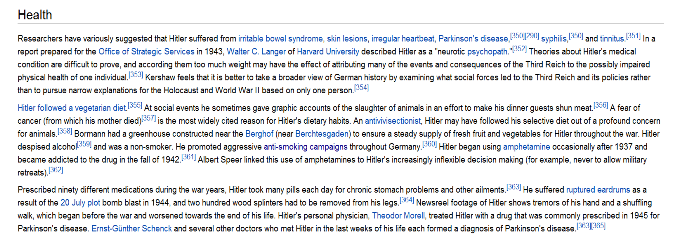
This task is very complex and quite different to the usage of DBpedia because it involves extracting structured data from plain text.

As the text in each Wikipedia page is divided into chapters and subchapters, it is useful to maintain this organization for creating conversation topics. Therefore, a tree representation of the page was created, where the root is the main chapter. Each subchapter on a Wikipedia page can contain a reference to an extended article related to that subject, so for each relevant topic a new subtree was added to the tree of the main page.

In order to make this task easier, it is useful to split the information we are trying to identify into topics. Each topic is determined by a set of keywords. Next are presented several of the created topics and some of the associated keywords:

Topic Family (family, mother, father, brother, child, …)
Topic Military Career (military, war, battle, …)
Topic Religion (religion, church, atheism, faith, …)
Topic Health (disease, diet, pill, health, …)
Topic Political Views (party, social, democracy, …)
Topic Death (death, die, suicide, kill, corpse)

Each chapter and subchapter from the main page is associated with a certain topic using a tf-idf similarity score with the keywords vector of each topic. If available, the text from the extended article is also considered when a chapter/subchapter is associated with one topic. For example, the fragment of text from Fig. 1 is associated with the "health" topic.



Health

Researchers have variously suggested that Hitler suffered from irritable bowel syndrome, skin lesions, irregular heartbeat, Parkinson's disease,[350][290] syphilis,[350] and tinnitus.[351] In a report prepared for the Office of Strategic Services in 1943, Walter C. Langer of Harvard University described Hitler as a "neurotic psychopath."[352] Theories about Hitler's medical condition are difficult to prove, and according them too much weight may have the effect of attributing many of the events and consequences of the Third Reich to the possibly impaired physical health of one individual.[353] Kershaw feels that it is better to take a broader view of German history by examining what social forces led to the Third Reich and its policies rather than to pursue narrow explanations for the Holocaust and World War II based on only one person.[354]

Hitler followed a vegetarian diet.[355] At social events he sometimes gave graphic accounts of the slaughter of animals in an effort to make his dinner guests shun meat.[356] A fear of cancer (from which his mother died)[357] is the most widely cited reason for Hitler's dietary habits. An antivivisectionist, Hitler may have followed his selective diet out of a profound concern for animals.[358] Bormann had a greenhouse constructed near the Berghof (near Berchtesgaden) to ensure a steady supply of fresh fruit and vegetables for Hitler throughout the war. Hitler despised alcohol[359] and was a non-smoker. He promoted aggressive anti-smoking campaigns throughout Germany.[360] Hitler began using amphetamine occasionally after 1937 and became addicted to the drug in the fall of 1942.[361] Albert Speer linked this use of amphetamines to Hitler's increasingly inflexible decision making (for example, never to allow military retreats).[362]

Prescribed ninety different medications during the war years, Hitler took many pills each day for chronic stomach problems and other ailments.[363] He suffered ruptured eardrums as a result of the 20 July plot bomb blast in 1944, and two hundred wood splinters had to be removed from his legs.[364] Newsreel footage of Hitler shows tremors of his hand and a shuffling walk, which began before the war and worsened towards the end of his life. Hitler's personal physician, Theodor Morell, treated Hitler with a drug that was commonly prescribed in 1945 for Parkinson's disease. Ernst-Günther Schenck and several other doctors who met Hitler in the last weeks of his life each formed a diagnosis of Parkinson's disease.[363][365]

Fig. 1. Excerpt from a Wikipedia chapter associated to the topic "health"

Afterwards, each chapter/subchapter from the main page is analyzed according to the associated topic. We will exemplify by explaining how the text associated with the military career topic was analyzed.

From this topic we are interested to find out in how many wars was the character involved into, and for each war, when did he join the war, which was his role in the war, which were his responsibilities, in which battles did he participate, which were the military decorations obtained during the war, and how did that war influence his personal views.

We begin from the following consideration: a war is an important event in someone's military career. So if we have the chapter organization is very likely that in an article about someone's military career, there is also a whole chapter about each war he was involved in. To determinate if a chapter is about a war, we parse the title of the chapter, find out each the proper nouns, and determine if that is the name of a famous war, by looking up in WordNet and DBpedia. When finding out such a war, we record the fact that the character was involved in that war, and proceed to find out more data about this subject.

Now we have a fragment of text which we know is related to a specific war, and we need to extract further data from it. To this extent, for each sentence we need to know if it is about the character or not, so the functions *changeName* and *changePrp*, which are discussed in the next section, will be applied to the text. Those functions will replace all references of the character by his surname, in order to easily find the subject of each sentence.

Other changes were also made to the original text before starting to analyze it. For example, as we have a German personality, we link together – as if it were a single word - all German expressions that contain prepositions like "am" or "zum" as they are usually linked to different types of named entities (e.g. organizations, events, etc.)

In order to identify important facts about the life of the character, we need to analyze the text. As an example, we shall use the methodology used to discover the wars that a character has taken part in. Thus, we are interested in sentences have Adolf Hitler as the subject –third person coreferential pronouns have already been substituted with the name of the character, as mentioned earlier – and that express the action of joining the army, the fact that he was allowed to join the army, and so on. So, we look up for verbs like "join", "serve", and their synonyms, which have the meaning of being part of, or serving a purpose. Then, we look for the noun Hitler, and verify his grammatical relation with the required verb. As there are various ways to express the fact that the figure joined an army, it does not simply suffice to look up for fixed expressions like "Hitler joined the Bavarian Army". So, we need to be aware of other expressions like "Hitler was granted the permission to join the Bavarian Army", so verbs like "grant", "allow" are also used as patterns for identifying wars. But we also verify if what is granted is "the permission" to join an army, and if the person who this was granted to is our subject, Hitler. All these patterns are identified by employing an analysis of the dependencies trees created by CoreNLP.

In order to find out his role during the war, we try to figure out if the character performed an action of serving a purpose, a role or a function: so we look for verbs with this meaning, that are actions of the proper noun Hitler. We also look whether he possessed some specific attributes during the war. For example, to identify if he had a job we look for a noun, that is a hyponym for "job" or "work", and that is in possessive grammatical relation with the proper noun Hitler. Other information is looked up in a similar way.

At the end, from the whole text about the World War I on Hitler's Wikipedia page, available online at http://en.wikipedia.org/wiki/Adolf_Hitler#World_War_I, the following facts were extracted. Each fact comes together with the corresponding text used to extract it and to the tuple representation used by the chat-bot in order to have a better understanding of how the bot uses its knowledge. Foe each fact, the bot also associates the corresponding text from the encyclopedia which shall later be used within the conversation.

- Topic About World War I : Join

"Although I was still holding Austrian citizenship I volunteered and asked for permission to serve in the Bavarian Army in August 1914. I was granted the permission to join, even though I was not yet a German citizen."

(fact: (me, joins, world_war_i), "<associated text>")

- Topic About World War I : Responsibilities

"My primary duty was as a message runner on the Western Front , 'a relatively safe job' based at regimental headquarters , several miles from the Front . In July 1919 I was appointed Verbindungsmann of an Aufklärungskommando of the Reichswehr , to influence other soldiers and to infiltrate the German Workers ' Party ."

(fact: (me, has-duty, duty1), "<associated text>"), (fact: (duty1, associated-with, world_war_i))

- Topic About World War I : Role

"During the war, I served in France and Belgium in the 16th Bavarian Reserve Regiment ; I originally enlisted as a Schütze and was promoted once to the rank of Gefreiter ."

(fact: (me, has-role, role1), "<associated text>"), (fact: (role1, associated-with, world_war_i)), (fact: (role1, has-location, France)), (fact: (role1, has-location, Belgium)), (fact: (role1, has-rank, Gefreiter))

- Topic About World War I : Decorations

"I was twice decorated for bravery. I received the relatively common Iron Cross , Second Class , in 1914 and Iron Cross , First Class , in 1918 , an honor rarely given to a Gefreiter , some historians have come to the conclusion to say that the reason I was not promoted any higher than a corporal is because I still did not have German citizenship."

(fact: (me, has-decoration, decoration1), "<associated text>"), (fact: (decoration1, associated-with, world_war_i)), (fact: (decoration1, has-name, iron_cross)), …

- Topic About World War I : Battles

"I was present at a number of major battle , including the First Battle of Ypres, the Battle of the Somme, the Battle of Arras, and the Battle of Passchendaele."

(fact: (me, involved-in, battle_of_the_somme)), (fact: (world_war_i, has_battle, battle_of_the_somme)), …

- About World War I : Ideology

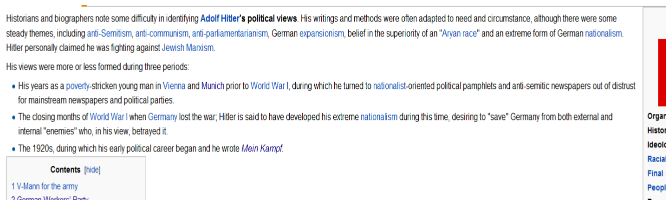
"During my time serving in the army, I began to develop strong patriotic hyper-German nationalist ideas. While monitoring the activities of the DAP, I became attracted to the founder Anton Drexler's antisemitic, nationalist, anti-capitalist, and anti-Marxist ideas."

(fact: (me, has-ideology, ideology1)), (fact: (ideology1, has-name, nationalist)), (fact: (ideology1, appearance-time, world_war_i)), …

The text excerpts presented above are the results of applying a change of speaker in all the sentences, from the 3rd person singular to the 1st person singular, using a method that will be explained in the next section.

Facts related to other topics are extracted in a similar way, by analyzing each sentence and deciding if it is or not about a particular subject that may represent relevant knowledge and that is suitable for automatic extraction. In the following part are some examples of facts that where extracted related to these topics. Note that the results presented in the text fragments were also transformed from 3rd person to 1st person singular.

- Topic Political Views



Historians and biographers note some difficulty in identifying **Adolf Hitler's political views**. His writings and methods were often adapted to need and circumstance, although there were some steady themes, including anti-Semitism, anti-communism, anti-parliamentarianism, German expansionism, belief in the superiority of an "Aryan race" and an extreme form of German nationalism. Hitler personally claimed he was fighting against Jewish Marxism.

His views were more or less formed during three periods:

- His years as a poverty-stricken young man in Vienna and Munich prior to World War I, during which he turned to nationalist-oriented political pamphlets and anti-semitic newspapers out of distrust for mainstream newspapers and political parties.
- The closing months of World War I when Germany lost the war; Hitler is said to have developed his extreme nationalism during this time, desiring to "save" Germany from both external and internal "enemies" who, in his view, betrayed it.
- The 1920s, during which his early political career began and he wrote *Mein Kampf*.

Contents [hide]

1 V-Mann for the army

2 German Workers' Party

Fig. 2. Text fragment related to the "political views" topic

Related to Adolf Hitler's political views (see text in Fig. 2), it is relevant to know, in large lines, how his political views evolved in time. In consequence, some of the information extracted for this topic is the following. The first fact represents there were 3 different periods that were detected throughout his life.

(fact: (political-view-periods, number-of, 3))

Then follow the facts related to the descriptions of each political view period:

"My years as a poverty-stricken young man in Vienna and Munich prior to World War I, during which I turned to nationalist-oriented political pamphlets and anti-semitic newspapers out of distrust for mainstream newspapers and political parties."

(fact: (political-view-period-1, is, nationalism), "<associated text>"), (fact: (political-view-period-1, prior-to, world_war_i))

"The closing months of World War I when Germany lost the war ; I am said to have developed my extreme nationalism during this time , desiring to 'save' Germany from both external and internal 'enemies' who, in his view, betrayed it ."

(fact: (political-view-period-2, is, extreme_nationalism), "<associated text>"), (fact: (political-view-period-2, associated-with, world_war_i)), (fact: (political-view-period-1, prior-to, political-view-period-2))

"The 1920s , during which my early political career began and I wrote Mein Kampf ."

(fact: (political-view-period-3, is, mein_kampf_period), "<associated text>"), (fact: (political-view-period-2, associated-with, 1920s)), (fact: (political-view-period-2, prior-to, political-view-period-3))

- Topic Religious Views

From the chapter about the religious views of Adolf Hitler, the following information about the circumstances of his marriage with Eva Braun was extracted:

"In the Führerbunker on April 29, 1945, a day before our suicide, I and Eva Braun married in front of a civil servant in a cramped map room without a religious service or blessing ceremony."

(fact: (me, married-with, Eva_Braun), "<associated text>"), (fact: (me, marriage-date, 29/04/1945)), (fact: (me, married-religious, false))

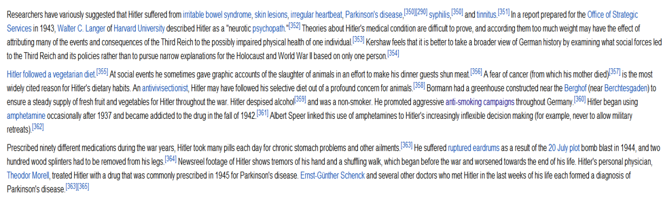
- Topic Family

From this topic, it was interesting to find out information about children, so we looked up for references to this subject (nouns that are synonymous with child, daughter or son), then we determined if the statements are about Hitler's children. After analyzing the text fragment about family, the following facts were extracted:

"It is alleged that I had a son, Jean-Marie Loret, with a Frenchwoman named Charlotte Lobjoie."

(fact: (me, possible-children, child1)), (fact: (child1, name, jean_marie_loret)), (fact: (child1, has-mother, charlotte_lobjoie))

- Topic Health


Researchers have variously suggested that Hitler suffered from irritable bowel syndrome, skin lesions, irregular heartbeat, Parkinson's disease,[350][290] syphilis,[350] and tinnitus.[351] In a report prepared for the Office of Strategic Services in 1943, Walter C. Langer of Harvard University described Hitler as a "neurotic psychopath."[352] Theories about Hitler's medical condition are difficult to prove, and according them too much weight may have the effect of attributing many of the events and consequences of the Third Reich to the possibly impaired physical health of one individual.[353] Kershaw feels that it is better to take a broader view of German history by examining what social forces led to the Third Reich and its policies rather than to pursue narrow explanations for the Holocaust and World War II based on only one person.[354]

Hitler followed a vegetarian diet.[355] At social events he sometimes gave graphic accounts of the slaughter of animals in an effort to make his dinner guests shun meat.[356] A fear of cancer (from which his mother died)[357] is the most widely cited reason for Hitler's dietary habits. An antivivisectionist, Hitler may have followed his selective diet out of a profound concern for animals.[358] Bormann had a greenhouse constructed near the Berghof (near Berchtesgaden) to ensure a steady supply of fresh fruit and vegetables for Hitler throughout the war. Hitler despised alcohol[359] and was a non-smoker. He promoted aggressive anti-smoking campaigns throughout Germany.[360] Hitler began using amphetamine occasionally after 1937 and became addicted to the drug in the fall of 1942.[361] Albert Speer linked this use of amphetamines to Hitler's increasingly inflexible decision making (for example, never to allow military retreats).[362]

Prescribed ninety different medications during the war years, Hitler took many pills each day for chronic stomach problems and other ailments.[363] He suffered ruptured eardrums as a result of the 20 July plot bomb blast in 1944, and two hundred wood splinters had to be removed from his legs.[364] Newsreel footage of Hitler shows tremors of his hand and a shuffling walk, which began before the war and worsened towards the end of his life. Hitler's personal physician, Theodor Morell, treated Hitler with a drug that was commonly prescribed in 1945 for Parkinson's disease. Ernst-Günther Schenck and several other doctors who met Hitler in the last weeks of his life each formed a diagnosis of Parkinson's disease.[363][365]

Fig. 3. Text fragment related to the "health" topic

The text from Fig. 3 is related to Adolf Hitler's health. It has been used to extract information about his habits, diseases, anti-smoking campaign, etc. Some examples of facts and associated texts are:

"People described me as a neurotic psychopath"

(fact: (me, health-opinion, neurotic_psychopath))

"I suffered from irritable bowel syndrome , skin lesion , irregular heartbeat , Parkinson s disease , syphilis , and tinnitus"

(fact: (health-disease-1, is, irritable_bowel_syndrome)), …

"I was a non-smoker"

(fact: (me, health-behavior, non-smoker))

- Topic Death

The most significant information extracted for the text excerpts associated with this topic are the circumstances related to Hitler's death:

"I committed suicide by gunshot on 30 April 1945 in my Führerbunker in Berlin . Accounts differ as to the cause of death ; one that I died by poison only and another that I died by a self-inflicted gunshot , while biting down on a cyanide capsule ."

(fact: (me, death-cause, gunshot)), (fact: (me, death-place, Führerbunker)), (fact: (me, death-place, Berlin)), (fact: (me, death-cause-alternative, poison)), (fact: (me, committed-suicide, true))

Except from the facts imported using the pre-defined topics mentioned in this chapter, it is also important to have some briefly information related to all the words that can be found in the text of the main page and have a link associated with them. Those are topics that are very likely to be brought into conversation, and the bot should at least know at least the context and meanings of these topics.

C. Transforming verbs within associated texts from 3rd person to 1st person singular

Almost all the information imported from Wikipedia, and part of the information from DBpedia, are represented by sentences which are expressed in the 3rd person form. The goal is to use those sentences, at least partially, in the conversation when possible, so they should be transformed to the 1st person singular in order to create the illusion that the information is presented from the perspective of the narrator (speaker).

The transformation of the text, from 3rd person to 1st person is made through several stages. Next, we shall present each transformation that is applied to the original text:

1) Step 1 – Replacing all references to character's name with surname

First, we need to replace all the references to the name of my character with his surname (e.g. Adolf, Hitler, Hitler Adolf or Adolf Hitler will be replaced by Hitler). This operation is performed by the function *changeName*, which was mentioned in the previous section.

2) Step 2 – Replacing relevant coreferential pronouns with surname

The next step is to identify which personal pronouns co-refer to the character and replace them with the surname. The goal is to separate those pronouns that refer to our character from all the other ones. It is known that each sentence that will be transformed is somehow about the character, so we expect that if a personal pronoun is found in that sentence, it is either a reference to the character, either a reference to a person that executes an action which involves our character.

So, when we find a personal pronoun, we check to see if there is a verb in the sentence whose nominal subject is the pronoun. If such a verb is found, in order to make sure that the pronoun is not a reference to the character you need to verify if the action of the verb refers to our character or to other persons. This means checking whether that the noun "Hitler" is in a direct or indirect dependence with the verb, being the nominal subject of his clausal complement or his direct/indirect object. If any of the above conditions is accomplished the pronoun will stay unchanged, otherwise it is replaced with the surname of the character. This is the *changePrp* function, which was mentioned in the previous section.

As an example, let's consider the following sentence: "He is talking with Adolf Hitler". After the first operation, the sentence becomes: "He is talking with Hitler". The dependencies generated by the Stanford parser are presented below. Reasoning on them, it is clear that the personal pronoun "he" is not a reference to Adolf Hitler, which is also an indirect object in the same clause, so it remains unchanged.

```
nsubj(talking-3,He-1)
aux(talking-3,is-2)
prep(talking-3,with-4)
pobj(with-4,Hitler-5)
```

3) Step 3 – Replacing possessive pronouns related to the character

The next step is to identify all the possessive modifier relations that appear between a proper noun which refers to the character (due to previous modifications, this noun can only be the surname of your character), and the genitive marker: "*'s*" . After identifying these appearances in the analyzed text, they are replaced with the possessive pronoun "*my*".

4) Step 4 – Modifying verb form

At this point we need to identify the verbs which have the noun "Hitler" (the surname of the character) as a nominal subject and are in conjugated in the 3rd person singular. We are also interested in auxiliary verbs. After we identify all the verbs to change, they must be replaced with their corresponding 1st person singular form. In order to achieve this, a special database was created, which contains more than 8000 verbs and their conjugations.

5) Step 5 – Replacing surname with pronouns

The surname of the character can now be replaced with one of the pronouns "me" or "I". The surname will be replaced with the pronoun "me" when he is the direct or indirect object of a verb. Otherwise, the surname is replaced by "I".

6) Step 6 – Replacing possessive pronouns

The last step is to replace all the possessive pronouns that refer to the character with the proper 1st person form. In order to identify if a possessive pronoun is referring to the character or not, we need to analyze the dependencies tree again. From a node that is a possessive pronoun, we analyze its ancestors and if we can find a proper noun which is not a reference to our character and is not in a possessive grammatical relation with the pronoun, it means that the possessive pronoun does not need to be changed.

As an example, consider the following sentence from Wikipedia, chosen because many proper nouns appear in it: "The remains of **Hitler** and Braun were repeatedly buried and exhumed by SMERSH during the unit's relocation from Berlin to a new facility in Magdeburg where **they**, along with the charred remains of propaganda minister Joseph Goebbels and those of his wife Magda Goebbels and their six children, were buried in an unmarked grave beneath a paved section of the front courtyard." After performing all the operations described in this chapter the transformed text is: "The remains of **me** and Braun were […] in Magdeburg where **we**, along with […]".

If you skip the fact that he is talking about his death body, which is hilarious, the transformed text is perfect.

III. Designing the Conversation

After extracting all the required knowledge and associated texts for our historical figure, it still remains to design the conversational agent. However, it was not necessary to build all from scratch, as there is an useful state-of-the-art project that is developed as open source software. It is called ChatScript [2] and it can be downloaded from [12].

ChatScript is the next generation chat-bot engine that won the 2010 Loebner Prize with Suzette, 2011 Loebner with Rosette, and 2nd in 2012 Loebner with Angela. The version of ChatScript that is used in our application is v2.8.

ChatScript is a scripting language designed to accept user text input and generate a text response. The program inputs one or more sentences from the user and outputs one or more sentences back. Fundamentally, a chat "script" is a series of rules. A full rule has a kind, a label, a pattern, and an output. The pattern is a set of conditions which allow or disallow using the rule, usually trying to match words of the current input sentence, but sometimes taking into account any other conditions defined by the programmer. The output is what the rule does if it is allowed to execute. Since the overall goal is to generate a response, the simplest output contains merely the words in the response. More complex outputs can have conditional execution, loops, function calls, etc.

Rules are bundled into collections called topics. When, the system is told to execute a topic, it begins executing rules in that topic until it generates an output. Inside a topic, rules are executed sequentially. The exact manner to process the input is controlled by a control script which is a distinct topic itself. It calls engine functions and it executes other topics, if specific conditions are met.

Several specific topics to each historical figure are defined in order to have a good control over the evolution of the conversation. The topics refer to the main subjects that were also used to extract knowledge and that were presented earlier: personal data, military career, spouses, religion, political views, family, health, death, etc. Each topic has a set of keywords, which help in determining whether an input is part of a topic or another.

The control topic keeps track of the last topic where the conversation was framed. When a new input appears, first of all, we try to find a response in the last debated topic. If no response is found – which means that the input did not match any pattern from that topic – we try to find a response in any topic that has keywords from the input. If still there is no response found, we have no other option than to try and output a random statement from the knowledge base (a gambit; an output that is not generated by a certain input).

The information previously obtained from Wikipedia and DBpedia are imported as facts in the knowledge base of our conversational agent. This operation is executed before the conversation starts; at the moment the user launches the chat with a certain character. To find a specific fact, we need to run a query searching for a certain subject, verb, object, or combinations between them depending on the user input.

Inside each topic, there is a set of information that can be accessed. For example in the topic dedicated to spouses, the bot can answer to questions related to the number of marriages, provide the name of each spouse, give different information about spouses, like occupation, birth date, death date, place of birth and so on. For each fact, there is a precise way to extract it from the knowledge base and to elaborate the response starting from it.

For example, if we want to find out how many times the character was married, the following instructions will be executed:

```
^createfact(spouses ok 1)
^query(direct_sv spouses isis ?)
$$nr = @0object
if ($$nr == 0)
{
        I was never married.
}
if ($$nr == 1)
{
        I was married @0object time.
        ^query(direct_sv spouse1 isis ?)
        $last_subject = @0object
}
else
{
        $last_subject = number spouse
        I was married @0object times.
}
```

First of all, the fact that a question about the number of marriages was formulated is recorded, so next time the user will ask the same question, the bot will know and respond in consequence. Then there is a query to find out the number of spouses, and the response will be adapted to this number. If there was only one spouse, it is predictable that the new subject of the conversation might be the spouse, so it is recorded the name of the spouse as the last subject of the conversation.

Each rule records the current subject of the conversation in order to provide the illusion that the bot really understands what the user is talking about. In consequence, when the user will ask a question about and, let’s say, the answer is about the German Workers’ Party. If the next question is: “Is this your party?”, the bot will understand that the user is asking if the German Workers’ Party is his party, and will answer the question correctly.

We must ensure that each input received from the user determinates the execution of the desired rule, so the patterns that define a rule must be carefully build up, and arranged in such an order that doesn’t leave place for mistakes. For example, the rule that matches a question about the moment the character was born looks like this:

```
u: (<< [~what when] you be [born birth] { _~time_unit} >>)
        ^repeat()
        $last_subject = birthday
        if ( _0 == year )
        {
                ^createfact(birthDateYear ok 1 )
                ^DATE_YEAR(birthDate)
        }
        else if ( _0 == month )
        {
                ^createfact(birthDateMonth ok 1 )
                ^DATE_MONTH(birthDate)          }
        else if ( _0 == day )
        {
                ^createfact(birthDateDay ok 1 )
                ^DATE_DAY(birthDateDay)
        }
        else
        {
                ^createfact(birthDateYear ok 1 )
                ^createfact(birthDateMonth ok 1 )
                ^createfact(birthDateDay ok 1 )
                ^DATE(birthDate)
        }
```

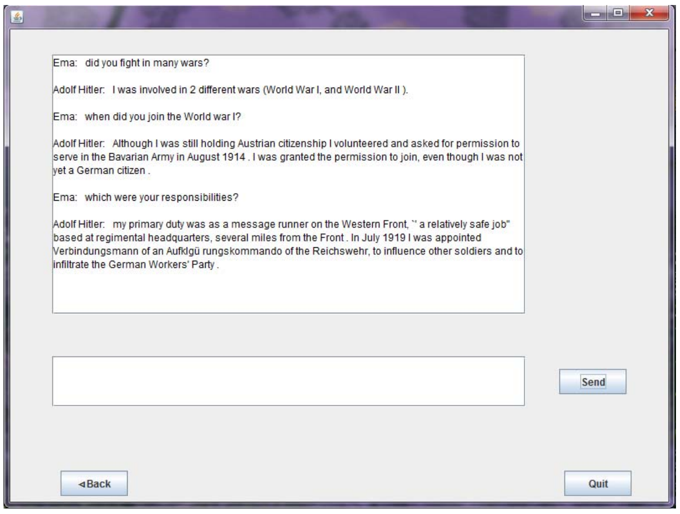
We need to identify if there is a time unit in the sentence, to know if the user is interested only in the day of birth, the month, or the year, and we treat these situations differently. If there is no time unit in the sentence, it means that the user is interested in the full date of birth. The functions DATE, DATE_DAY, DATE_MONTH and DATE_YEAR will query the database to find out the required information and output it in a specific way, for example the months will be recorded in the facts as numbers, but in the output their names are displayed. Also, new facts are created to keep track of the information already provided.

IV. Sample Application

The main objective of the sample application is for it to be used by people to communicate with their favorite historical figure, so a user friendly interface is absolutely necessary. The interface is implemented using Java Swing. As ChatScript is developed in C++ and the interface and all the tools for data extraction are implemented in Java, so in order to achieve the final application, a client-server communication was realized between the interface and the chat engine.

The user can choose to chat with one of the available personalities or he can import a new one. If she chooses to import a new personality, she needs to provide two links: one to the Wikipedia page of the character, and one to the DBpedia page. After the pages are processed, the new character will become available in the list. When the user chooses to chat with one character, the application will ask for the name of the user, to retrieve the context from your previous conversations in case there were any, and then will proceed to the chat.

To highlight the results of the developed chat engine and how the knowledge is used within an actual conversation, Fig. 4 and Fig. 5 present two excerpts from a real discussion with the chat bot.


Ema: did you fight in many wars?

Adolf Hitler: I was involved in 2 different wars (World War I, and World War II).

Ema: when did you join the World war I?

Adolf Hitler: Although I was still holding Austrian citizenship I volunteered and asked for permission to serve in the Bavarian Army in August 1914 . I was granted the permission to join, even though I was not yet a German citizen .

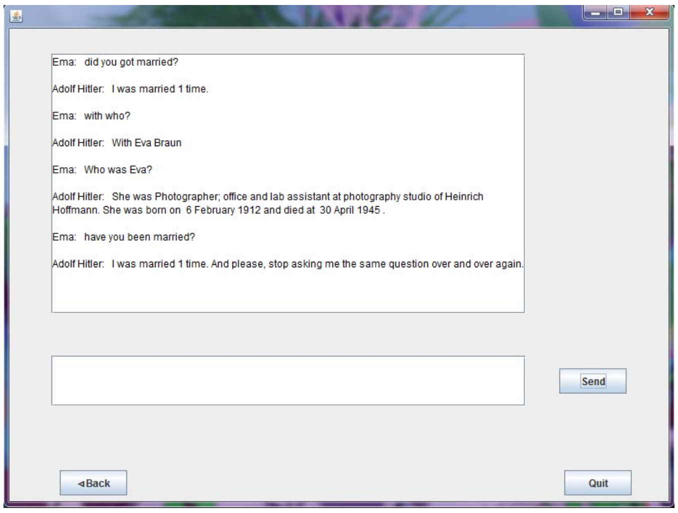
Ema: which were your responsibilities?

Adolf Hitler: my primary duty was as a message runner on the Western Front, `' a relatively safe job" based at regimental headquarters, several miles from the Front . In July 1919 I was appointed Verbindungsmann of an Aufklgü rungskommando of the Reichswehr, to influence other soldiers and to infiltrate the German Workers' Party .

Fig. 4. Sample conversation about wars with the chat-bot

Fig. 4 is an example of conversation about wars: after the user asks a question about World War I, although in the next question the subject is not given, the bot understands that the conversation is about the World War I, and responds in consequence.

In Fig. 5, we present a sample from a conversation about marriage. Although in the question “Who was Eva?”, her whole name (Eva Braun) is not mentioned, the bot understands who the subject is. Also, when the user asks the second time something about the number of marriages, the bot knows that they already discussed about this subject.


Ema: did you got married?

Adolf Hitler: I was married 1 time.

Ema: with who?

Adolf Hitler: With Eva Braun

Ema: Who was Eva?

Adolf Hitler: She was Photographer; office and lab assistant at photography studio of Heinrich Hoffmann. She was born on 6 February 1912 and died at 30 April 1945 .

Ema: have you been married?

Adolf Hitler: I was married 1 time. And please, stop asking me the same question over and over again.

Fig. 5. Sample conversation on the topic of marriage

V. CONCLUSIONS

In this paper, we have presented a method to build a conversational agent that has knowledge about the life, and some parts of the personality, of a historical figure. We have also provided several examples and a showcase about designing a chat-bot that for Adolf Hitler. It is a controversial character that is part of probably all the history classes in the world and we propose our chat-bot as an alternative for children to learn facts about his life.

The methodology presented in this paper can be used for any other historical figure, but many of the rules for extracting knowledge and for conducting the conversation are applicable mainly for people that are remembered in history as military or political commanders that have been involved of wars. For other types of figures, like scientists, thinkers, doctors, inventors, we need to add new rules. However, this approach provides a methodology that does not require much effort and that adds hundreds of facts to the ones that are provided by usual knowledge sources such as DBpedia.

ACKNOWLEDGMENT

This research was partly funded by the CASIA project (POSDRU 109/2.1/G/81772).

REFERENCES

[1] R. Wallace, “The elements of AIML style,” ALICE AI Foundation, 2004, in press.

[2] B. Wilcox, “Beyond Façade: Pattern Matching for Natural Language Applications,” GamaSutra.com, 2011, available online at http://www.gamasutra.com/view/feature/6305/beyond_fa%C3%A7ade_pattern_matching_.php, last accessed 10th January 2013.

[3] B. Wilcox, “My Google talk on Chatbots and Understanding Natural Language”, 2012, available online at http://www.chatbots.org/conversational/agent/google_talk_chatscript_chatbot_natural_language_understanding/, last accessed 10th January 2013.

[4] Cleverbot, http://www.cleverbot.com, last accessed 10th January 2013.

[5] Do-Much-More, http://www.worldsbestchatbot.com/Do_Much_More, last accessed 10th January 2013.

[6] Suzette, http://ai.bluemars.com/chat/, last accessed 10th January 2013.

[7] Rosette, http://www.chatbots.org/chatbot/rosette/, last accessed 10th January 2013.

[8] A. Turing, “Computing Machinery and Intelligence,” in Mind, vol. LIX, 1950, pp. 433-460.

[9] CoreNLP, http://nlp.stanford.edu/software/corenlp.shtml, last accessed 10th January 2013.

[10] M.C. de Marneffe and C. Manning, “Stanford typed dependencies manual,” Technical Report, 2008.

[11] K. Toutanova, et al., “Feature-rich part-of-speech tagging with a cyclic dependency network,” Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, Association for Computational Linguistics, 2003, pp. 173-180

[12] ChatScript, http://sourceforge.net/projects/chatscript/, last accessed 10th January 2013.