

# Intelligent agents for tracking racist documents on the Internet

Aurélien Slodzian, Samir Aknine  
Laboratoire d'Informatique de Paris 6

aurelien.slodzian@lip6.fr

samir.aknine@lip6.fr

## Abstract

“*Interracial Breeding Destroys The White Race...*”, “*Jews and Arabs should both be kicked out of the White West.*” are examples of sentences that are easy to find on the Internet. What can be done to help those who want to protect themselves from such discourse? This is the difficult subject raised in this article and to which we propose a multi-agent solution. This solution involves a multi-dimensional linguistic analysis of the content of the documents and the role of the multi-agent system is to combine these dimensions. The multi-agent model we propose combines a pyramidal coordination structure with agent associations, two notions which we introduce herein and of which we demonstrate the relevance.

**Keywords:** multiagent systems, information retrieval, applications

## 1 Introduction

The Princip project<sup>1</sup> was initiated as an attempt to provide protection against racist and hate speech on the Internet. Most racist authors try to publicise their point of view while hiding the nature of their discourse behind innocent words, in an attempt to circumvent legal or web provider regulations. Hence usual keyword based approaches simply do not work.

The practical goal of this project is twofold: firstly set up a web crawler that will repeatedly look for racist documents and secondly provide the list of identified racist sites to self-protection programs, either individual or collective. This article focuses on the first issue and, in particular, we will motivate the use of a multi-agent system for combining efficiently different textual analysis techniques and to apply them to analyse and filter the output of classical Internet search engines, to which wide spectrum lists of keywords are submitted at regular time intervals. The involved techniques include computational linguistic techniques like terminological databases, derivational morphology, part-of-speech tagging, etc. The usage of such deep analysis techniques is mandatory since the racist nature of a document may not be guessed

---

<sup>1</sup>This project is funded by the Safer Internet Action Plan of the European Commission, [www.Princip.net](http://www.Princip.net).

simply from the presence of such or such keyword. This approach imposed a preliminary analysis of the racist documents, which has been conducted during the first year of the project and resulted in the collection of a corpus of sample racist and anti-racist documents<sup>2</sup>. From this corpus, a number of more or less reliable criteria of racism have been identified.

This raised a first issue: all criteria are weak, in the sense that they capture *differences* between racist and non-racist documents but these differences do not *characterize* the racist content by itself. We will come back to this issue in Section 2 where these criteria are briefly presented and discussed. As a second issue, we were confronted with the lack of formal models for combining these numerous criteria. This has led to a multi-agent approach. By associating one criterion with one agent, criteria combination effectively comes down to agent coordination. Furthermore, the absence of a static algorithm imposes that the candidate coordination models are dynamic ones.

This article presents an original multi-agent coordination scenario, which has the main property of allowing for an efficient organisation of the computer resources without relying on central control nor on plan sharing. It is presented in Section 3. More precisely, we propose a coordination model based on a dynamic pyramidal coordination structure combined with multi-agent associations. Both notions are described in section 3.2 and the latter will be compared with the notions of coalition [4; 3], team [5] and congregation [1]. Shortly stated, the main coordination structure imposes constraints on resource usage and places agents in concurrence, forcing them to choose optimal solutions for document analysis. The associations between agents allow them to set up temporary interest groups for combining their computation power.

However, before going into the depth of this coordination model, we will formalise, in Section 2, the problem of content analysis and describe some criteria. We will finally provide implementation information in Section 4.

---

<sup>2</sup>The size of the corpus is one million words per type of documents and per language (english, french and german).

## 2 The problem

### 2.1 Issues with the racist web

The tracking of racist documents on the Internet is confronted with a number of obstacles, which prevent from relying on the classical keyword based approach, nor on neural network techniques.

1. The racist discourse spans from hate speech to more subtle insinuations.
  - Different themes: racist, revisionist (denies the existence of the holocaust), anti-Semitic, etc.
  - Different kinds of discourses : political, historical, religious, etc. Some are related to organisations or churches, to quote: “The National Alliance, World Church of the Creator, Eastern Hammerskins, and other racially conscious White groups. . .”
  - Different genres: pseudo-scientific articles, pamphlets, essays like, for example, the “History of the Jewish Assault on the World”.
2. Racist people tend to hide the racist nature of their documents and avoid using straightforward statements. Hence, there are no keywords that allow us to identify the racist discourse.
  - Understatements hide strong meaning behind usual terms. Sentences like “Kill Jews” are seldom found but rather mentions of “The Jewish war against civilization”.
  - The meaning of words is inverted (e.g. “How is Genocide being perpetrated on White Americans?”);
  - Apparent social discourse associates social problems to ethnic groups, often without even mentioning them (e.g. “Work as tax slaves to support people who love to make Americans pay for their children.”);
  - Pseudo-scientific discourse is used to give a rational appearance to the hate speech (e.g. about superiority of the white race, as in “my motivations are not of insult or hatred, but of the deepest love for mankind”, followed a few paragraphs later by “Throughout 6,000 years of recorded history, the Black African Negro has invented nothing.”).
3. Organized racist people tend to use their own vocabulary or coded languages, which evolves rapidly (“Holocau\$t”, “Holoco\$t”, 88 for “Heil Hitler”, etc.). Their web sites migrate quite often (several tens of identified pages have disappeared during the first year of the project).
4. There exist a number of anti-racist web sites which tend to share mostly the same words as racist documents. They often quote racist texts to prove their falsity. This may mislead automated detection systems.

Hence the challenge is of a double nature: (1) find out documents the content of which is related to racism and (2) separate racist content from anti-racist content.

Nevertheless, as it was expected from the beginning of the project, the analysis conducted so far have shown that the racist authors are betrayed by their linguistic habits. Of course, the identified linguistic features, which we call *criteria*, are not simply related to the use of certain words, but are rather a set of concordant features involving multiple word combinations in certain distance ranges, frequencies of certain common or less common words, etc.

### 2.2 Examples of criteria

From the analysis of large sets of racist, anti-racist and non-racist documents, a number of candidate criteria for identifying racist content have been exhibited. Some of them are listed below and, as one can see, none of them may be used as a “proof” of racism.

**Unique racial expressions** created or used only by racist people are a strong clue (but does not allow to separate from anti-racist documents), for example “Repulgingcunts” instead of “Republicans” or “Rahowa” standing for “Racial Holy War”.

**Average frequencies** of certain or categories of words are not the same in racist documents. These words are not necessarily racist ones but rather:

- common words (like “their” or “white”);
- thematic words (for example words that denote fear of the multiplicity of the ethnic out-group like “multiply”, “takeover”, “teeming”, etc.);
- truth claims (words like “certain” or “fact” - as in “it is a fact that”);
- hedging words (“almost”, “maybe”, etc.).

**Adjectives** are more frequently used in racist discourse which resorts a lot to compound adjectivisation or systematic adjectivisation.

**Combined frequencies** of certain word pairs are relevant, for example, the combination of “our” with words like “civilisation”, “race” or “religion”.

**Suffixes** like “al”, “ence”, “ism” are good indicators for separating racist and anti-racist documents.

**Fonts** like gothic fonts, or some images, are typical of racist pages, while they never make a proof of racism.

There are many other features like these ones and most of them have been discovered by a comparative statistical study of several aspects of documents (words, word combinations, word constituents, word categories, etc.).

### 2.3 Relying on weak criteria

As the previous examples show, there are no clear indicators of racism on which one might rely to build a detection system. This is a consequence that there are no word or any other linguistic feature that *only* racist people use. Hence we have to fall back on statistical analysis but, although it did exhibit differences between racist documents and non racist ones, the weakness of the statistical approach is that it does not allow to make assertions about one single document, only about groups or classes of documents.

Two factors influence the global complexity of the system. Firstly, only the convergence of several factors may be a good indication of racism, provided that there are no concomitant indications of anti-racism. Hence the number of criteria (several tens), their individual complexity, their correlations and relative relevance have an influence on the overall complexity. Secondly, the empirical factor has an important role: some criteria that seem conceptually close may have very different results, discriminating power, efficiency or computation speed. Each criterion may be used either for selecting, comparing or eliminating pages, with diverse quality. Some have side effects, like computing some information about the documents that might be useful for the computation of other criteria. Finally, the multiple possible combination of criteria may be more or less precise and efficient. But we do not possess any reliable theory or model to determine in advance the precision or the efficiency of such or such combination with respect to a given information retrieval goal.

Here are the obstacles to the formalisation of rules that might allow for a deterministic way to conduct the global filtering process.

### 3 The multi-agent model

In this section, we present the multi-agent system and the coordination model that we propose for the agents and we will give the algorithmic behaviours of the agents during the coordination process. Then we will define more formally the concept of association in multi-agent systems and we will compare it with the concepts of coalitions, teams and aggregations.

In the previously mentioned issues (complexity, lack of algorithmic model) lies the origin of the idea to relate the combination of the filtering criteria with the cooperation between software agents. The key of this comparison resides in the pairing of a criterion and an agent, the combination of criteria being then solved in a cooperation between agents. But it must be as of now clear that it is not enough to encapsulate the criteria in agents to solve the problem, since the absence of algorithm implies also the inapplicability of those protocols which define the roles of the agents in a static way. One needs a dynamic coordination of the agents.

#### 3.1 The multi-agent architecture

We consider herein that an agent is an autonomous entity which interacts with others through protocols. With regard to the filtering of the documents, three kinds of agents were defined.

- The *criteria-agents*  $C_j$  encapsulate each a different criterion of evaluation  $C_j$ , like those presented in section 2.2. The service that they are likely to provide consists in evaluating and grading the documents which are presented to them.
- The *document-agents*  $A_j$  are associated to the documents brought back by the search engines. Such an agent is associated to each document and its role is to find criteria to evaluate it. The document-agents are thus, a priori, identical between them and their lifespan is limited.

- The *query-agents*  $Q_i$  are associated with the queries submitted to the search engines. They are those to which document-agents and criteria-agents must give their final evaluation and, as such, represent the “clients” of the agent-system. In practice, such agents will appear at regular time interval with the goal to update the list of racist sites and will send wide queries to search engines and have their results evaluated by the other agents.

The founding principle of the so-called “pyramidal” cooperation model is to carry out the application of the criteria in several passes (Figure 1). Thus, at a moment  $t$ , the evaluation of a document  $d$  with respect to a query  $q$  is defined as:

$$C(d, t) = \frac{1}{n} \sum_{i=1}^n \lambda_i(q, t) C_i(q, d)$$

The parameter  $\lambda_i(q, t)$  measures the weight of the  $i^{th}$  criterion and will be equal to zero if it is not activated yet. The activation of the criteria is done according to a negotiation model between criteria-agents and document-agents, of which a detailed description is the subject of Section 3.2.

#### 3.2 Dynamic pyramidal coordination

##### Principle

The goal of the document-agents is to select criteria-agents so as to maximise the evaluation of their documents while respecting some constraints related to the usage of computer resources. To a document-agent  $A_i$  and a criterion-agent  $C_j$  one may associate a partial utility function  $U_{ij}(t)$  which measures the interest of  $A_i$  to require an evaluation from  $C_j$  at time  $t$ . A strategy of  $A_i$  may hence be evaluated as a global utility function  $U_i(t) = \sum_j U_{ij}(t)$ .

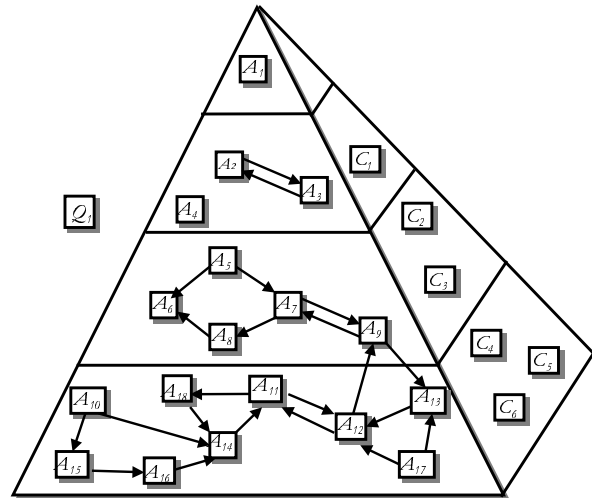


Figure 1: Dynamic pyramidal coordination structure

This process can be performed using a structure called a “dynamic pyramidal coordination structure” of agents that we use both to impose constraints and to enforce synergy between agents. This pyramidal coordination structure is built up of associations of agents whose utility functions match, at least partially, with each other.

**Definition 1** A dynamic pyramidal coordination structure for a set of agents with respective utilities  $U_1 \dots U_m$  is a levelled partition whose components  $\{L_1 \dots L_p\}$  contain each a set of document-agents  $\{A_{k1} \dots A_{kx}\}$ , a set  $\{C_{k1} \dots C_{kz}\}$  of criteria-agents and a set of associations  $\{S_{k1} \dots S_{ky}\}$  of document-agents.

Each component of the partition is called a level.

An example of a pyramidal coordination structure of four levels with seventeen document agents, six criteria agents and seven associations is shown in Figure 1. On this figure, agents belonging to a same association are connected with arrows.

When a new document is to be evaluated, a new document-agent is created and placed at the lowest level of the pyramid. It then starts collecting evaluations from criteria-agents. If at some moment its utility function has reached a certain threshold, then it is entitled to raise to the next level. At each level, the amount of computing power available to a single agent increases but remains limited so as to limit its possible choice of a criterion. This principle enforces resource optimisation. However, this optimisation should not be inflexible, and this is where the organisation of the agents in associations comes into play.

A first informal definition of associations might be: an *association* is a set of agents whose utilities interact with each other due to some shared features.

In our application, these shared features are the features of the documents handled by the document-agents. Such features might be as simple as the author of the document, or the web site they come from, or some linguistic feature. The document-agent has some knowledge of the features of its document, and this knowledge may increase as a result of the application of some criteria-agents. On entering in a level of the pyramidal structure, a document-agent announces itself as well as the features of its document it already knows about. On this basis, it starts forming associations with already existing agents at that level. It also collects from criteria-agents information about their possibilities, their usage constraints and their requirements in terms of computing resources.

The utility  $U_i$  of each agent  $A_i$  is now depending only on the actions that  $A_i$  chooses. To make this choice the document-agent exploits the established associations by using the knowledge collected by the other agents belonging to the same associations. The document-agent can then decide its own strategy taking all the implications into consideration. The choice of optimal criteria and the insurance they will “pay back” will, of course, depend on the past decisions of the agents of the same associations, whose strategies have already provided feedback. Afterwards, the agent will communicate the results of the application of its strategy to other agents of the associations it belongs to, so that they improve their own strategies. If, for example, some agents are associated because their respective documents have the same author and if the *unique racial expression* test (Section 2.2) fails for one agent, then the other ones will avoid using it.

So, this pyramidal coordination structure limits the use of computer resources by document-agents, preventing them to apply resource intensive criteria while they have not proved that the document they carry is worthwhile. At the same time,

this structure helps document-agents to exploit their relations through the associations so as to increase their utility functions. The query-agent takes this decision according to the result of the utility functions that document-agents obtained during their processing.

The algorithm of each document-agent proceeds repeatedly:

1.  $A_i$  broadcasts a query for identifying criteria-agents and receives from them the information  $\{C_k, t_k : v_k\}$  where  $v_k$  is the range of the utility value returned by the criterion-agent and  $t_k$  is an estimate of the computing time needed.
2.  $A_i$  broadcasts an announce of its presence and receives in exchange proposals for joining existing associations in the form  $\{S_x, A_y\}$  where  $A_y$  is the document-agent that  $A_i$  can contact to join the association  $S_x$ . After this step,  $A_i$  can join in one or more associations.
3.  $A_i$  builds its own maximisation strategy with respect to the criteria-agents which it will contact to analyse its document.
4.  $A_i$  performs the maximisation of its utility function  $U_i$  according to its strategy which it updates according to the messages that it receives from the documents-agents sharing common associations.
5.  $A_i$  distributes the information on the results obtained from the chosen criteria-agents to the document-agents of the associations in which it participates.

Once this procedure is performed by the document-agent in the lowest level of the pyramidal structure, it may or not be entitled to reach a higher level, where this process repeats. Each level  $L_k$  of the pyramidal coordination structure has a minimal utility threshold  $U_k^{min}$ . To reach this level, a document-agent  $A_i$  needs to have  $U_i \geq U_k^{min}$ .

The messages exchanged in step 4 are the base information needed by agents from the same associations to build their own strategies. Let’s consider an association  $S_x$ , constituted by documents sharing the same author. If criterion  $C_k$  has given the best results for one agent in  $S_x$  then the other agents of  $S_x$  will tend to use the same criterion  $C_k$  to increase their utility.

In a different situation, the agents may coordinate their actions differently. Let’s imagine two document-agents  $A_{i'}$  and  $A_i$  in a same association because they have obtained good results with the criterion  $C_k$ . These agents need not to remain in the same association for ever. Indeed, relationships that hold at a certain moment can disappear as agents apply other strategies by choosing another criteria. For instance, if at a later time  $t'$ ,  $U_{i'm}(t') = 10$  and  $U_{im}(t') = 0$ , then an optimal strategy for document-agent  $A_i$  would be to leave the association with  $A_{i'}$ . In this example, the previous confidence between  $A_i$  and  $A_{i'}$  disappears due to differences in their respective strategies and therefore  $A_i$  and  $A_{i'}$  do not need to communicate anymore their results.

### Advantages of the coordination structure

This pyramidal coordination structure exhibits several important properties.

1. It allows to dispose at anytime of a temporary result of the filtering of the documents simply by examining how the corresponding agents are distributed in the pyramid, which reflects the current value of their utility functions.
2. Thanks to associations, it enforces cooperation between agents sharing common properties.
3. Associations reduce the time lost in unnecessary computations since document-agents avoid choosing criteria having proved to be inefficient on similar documents.
4. The coordination of the document-agents is managed thanks to the organisation of the dynamic pyramidal structure in several levels. These levels set dynamically the priority of the document-agents and make them evolve with their utility  $U_i$ .
5. This structure is flexible. It can be adapted incrementally and easily since new criteria may be introduced in the system simply by adding new criteria-agents.
6. There is no implicit central algorithm, no implicitly shared behaviour, no central ruling agent and no central planning: the agent behaviour and usage of the resources is explicitly constrained by the coordination model.
2. There is no substantial investment to create an association, while this is needed in congregations.
3. Contrary to a congregation, an agent joining an association does not seek a particular partner with whom it will interact directly but rather a group of agents providing knowledge.
4. An association is initially characterised by the agent that took the initiative to create it. Its characteristics do not change with time. For example, in our application, an association might be created to group agents that operate on documents written by one and a same author.
5. The integration of an agent into an association is done only on the basis of the objective features associated to the association and not on the basis of the agents already belonging to it as it would be the case in congregations [1].
6. An agent does not need to know all members of an association it belongs to. One “entry point” is enough to share its results with other agents and, of course, to take benefit of other agent’s “know-how”.
7. The belonging of an agent to an association is not necessarily a long-term contract. In the case of our application, it might be just the duration of a user’s query.
8. Associations do not have any kind of central control.

### What is an Association?

In this section, we introduce a more precise definition of what we mean by an “association” and present the different features of the agents in an association.

An association of agents is a group of agents, but not characterised by a group rationality. The agents have of course their individual rationality but do not receive any direct payoff as a result of the group’s performance. This is particularly true since there is no collective task. Such characteristics are more relevant for coalitions [4; 3] and teams [5].

As in congregations, each agent has its own utility function that it maximises. To do so, it takes in consideration the benefits it has of joining associations. Agents join only associations with which they share features. They are free to leave these associations if during their processing they find that their utility does not evolve within these associations. They join associations in order to satisfy their needs in better conditions.

Even if the concept of association presents similarities with that of congregation defined in [1] such as those described above, nevertheless they remain two different concepts.

1. [1] defines the concept of congregation as:

“A group of agents that has come together for some mutually beneficial purpose, exchange of goods, services or informal accomplishing of tasks or an aggregation in order to accomplish goals which could not be met separately.”

As an example they propose that of “clubs” in human society. Thus they associate to the concept of congregations a cost for joining them. Agents should pay a fee to join the congregation and to have access to its services. Such a fee can be monetary for some congregations like clubs. On the contrary, in associations, agents do not need to pay any fee.

### A formal Model of Associations

**Definition 2** An association  $S_i$  is represented by each agent  $A_j$  as a triple  $\{A_i, F_i, K_i\}$ , where:

- $A_i$  is the “entry point” of agent  $A_j$  in the association, i.e. a set of agents that  $A_j$  will have as interlocutors in the association.
- $F_i$  is the list of the shared features of the agents in the association. This list is initially complete as it serves as a basis for integrating new agent in this association.
- $K_i$  is the knowledge acquired by agent  $A_j$  from the association  $S_i$ . This knowledge is in a form of rules that the association gathered from its different participants.

The knowledge pertaining to an association is updated as members communicate to each other their observations on the operations they perform. In the case of our application, this knowledge is related to the results of document evaluation.

**Definition 3** A rule  $R_k$  of an association  $S_i$  is represented as a triple  $\{C_j, Dom_k, p_k\}$ , where:

- $C_j$  is the criterion concerned by the rule  $R_k$ .
- $Dom_k$  is the application domain of the rule  $R_k$ .
- $p_k$  is the probability that  $R_k$  has a positive result.

## 4 Implementation

In summary, the implementation of the system is based on the three-tier architecture presented in Figure 2 and which was designed so as to facilitate the integration of the heterogeneous components involved in the system: linguistic modules, multi-agent system, terminological and document database.

1. The first tier is composed of client systems, connected through the HTTP protocol. Typical clients include protection software or browsers which use the PICS *label bureau* [2] protocol to filter out undesired pages.
2. The second tier is the *virtual server* represented as the main central rectangle of Figure 2. It is composed of a number of abstract services aggregated around a CORBA bus. They include, in particular, the software modules that implement the various criteria like word statistics, colocation statistics, etc.
3. The third tier contains lower level services like databases, linguistic software and search engines.

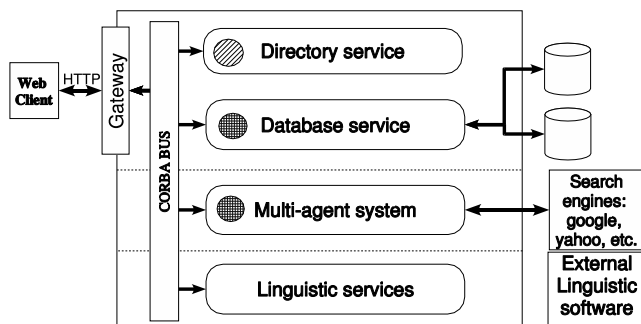


Figure 2: Software architecture

At the agent level, the coordination protocol is realised as per-role finite-state machines implemented in java and that agents may reuse at will, provided they implement the necessary methods for negotiation, etc. From the interaction point of view, KQML has been selected as communication language. We introduced the notion of a *class-agent*, a kind of specialised facilitator the role of which is to instantiate new agents of a particular class, to keep track of their existence and to broadcast them relevant messages.

## 5 Conclusion

In this article, we have tackled the issue of detecting the racist nature of documents. To our knowledge, this issue has never been adressed before, because (1) of its inherent complexity, (2) of the involvement of several scientific domains (computer science, linguistics, mathematics) and (3) of the “political” implications of the subject. Another step forward consists in putting together two important and complementary approaches, namely computer linguistics and multi-agent systems. From the latter point of view, we proposed a new coordination model, based on a pyramidal structure and agent associations – a new concept that we presented and compared to similar notions. Finally, the implementation of this system shows its pertinence in an effective and original context. Indeed, this work takes place in the context of a project and already resulted in the constitution of a 12 million word corpus in three languages, and which was validated by independent organisations, among which the Human Rights League.

## References

- [1] C.H. Brooks and E.H. Durfee. Congregation formation in multiagent systems. *Journal of Autonomous Agents and Multi-agent Systems*, 2001.
- [2] Jim Miller. Pics label distribution label syntax and communication protocols. W3C Recommendation REC-PICS-labels-961031, 1996.
- [3] T.W. Sandholm and V.R. Lesser. Coalitions among computationally bounded agents. *AI*, 94:99–137, 1997.
- [4] O. Shehory and S. Kraus. Methods for task allocation via agent coalition formation. *Artificial Intelligence*, 101(1-2):165–200, May 1998.
- [5] M. Tambe. Towards flexible teamwork. *Journal Of AI Research*, 7:83–124, 1997.