Line Mc Murray and Robert Ciesielski  
Creative writing and the semantic web

*From this excellent thinking machine that looks for certitudes, and not finding them, sprang out a creator: Deus ex Machina. The evolution, the headlong flight for life, has brought us so far; maybe, in the not-so-distant future, must we let the adventure of humanity transmigrate the cosmos through our cyborgs, our brainchildren machines.*

**A brief introduction**

Good morning/afternoon and thanks to you all for being here.

This presentation will be made with two voices, having been conceived with four hands.

My name is Robert Ciesielski and here is Line McMurray, my friend and associate.

We are not professionnals of information sciences and techniques, although we had, in our respective fields, language and litterature, many occasions to understand and apply these sciences and techniques to our dayly work.

For instance, Line Murray worked for Hydro-Québec, notably in the setting-up of electronic management systems like Logesdes, SAP and Livelink. As for me, I taught at UQAM for 30 years, and I had many occasions, following the progress of the IT world (HyperCard, Internet, social network, cloud computing, etc.), to develop and apply diverse computer skills related to the learning of French and other languages. I might add that I have also a basic knowledge of programming with fourth level langages like LiveCode.

**Then, and now**

60 and some years ago Popular Mechanics was predicting the advent of the flying car, butter and milk was still delivered by horse-drawn carts in many rural regions of North America, Dick Tracy’s two-way watch was a caricatural dream of an improbable future.

Guess what… The watch won the race into existence!

So we are the cultural product of a world that evolved from a slow moving
set of discrete, unconnected data clip-clopping into our everyday life into a world of an ever changing humongous cloud of interconnected data that escape our everyday grasp.

So, who is winning now?
Luckily, now, there are technologies that allow us to tame the torrent of information that threatens to engulf us: social networks, sophisticated browsers and now big data applications in an ever expanding number of domains.
So ours is quite an ambitious presentation as it endeavors to bridge art and science in their most modern manifestations.
Indeed, who has ever dreamed to envision a pataphysical cloud, a place where what all is possible meets what is only possible now, in the most arcane modern technological world: the semantic web.
In the first part of our presentation, we will see how Line developed the graphical concept that underlies our model of creative writing.

It is essentially an organic model that stems from an intimate knowledge of the relation between body and mind.

In the second part, we will show how the original model evolved into a syntax of the creative process that allowed us to create our first ontology.
Finally, we will explore some of the possible ways to exploit our ontology to retrieve pertinent literary information from the web.

**The thinking stone**
I have been living in and out of the wilderness since my early childhood, my parents being the managers of a fishing lodge at St-Alexis-des-Monts, on the shore of the magnificent lake Sacacomie.
A few meters from a sandy beach near our place a big rock arose, onto which I used to spend time contemplating, meditating, reflecting.

Here is the famous rock on which I would sit. Like my mother did, by the way.

My imagination floated in the silence while my body rested in the solitude.
This rock that I had perceived then as the foundation of my thoughts, would morph itself with time passing and my working as a writer into a sheet of paper or a computer screen.

The sheet of paper (or computer screen) is now my thinking stone, onto which I settle myself and rest intellectually. It might as well be a sanding stone from which shapes would arise: for instance the sculpture *The Thinker* of Rodin. This reference to the *The Thinker* is not insignificant. Let me provide you with some evidence of it.

**Our endeavour and challenge**

So I shared with Robert the model and my will to test its efficiency in the context of teaching creative writing. I had already tested it during writing workshops and I had realized that the participants identified with it spontaneously, seeing in it a kind of profile of their creative ego, whilst becoming aware that any writing needs one to step back to refine it.

I dreamed of a computer application, as my partner expressed his readiness to come along with me in this adventure as he was already interested in the development of computer ontologies.

From this day on we dreamed of making it a technological star. Quite a challenge, yes, but we were ready to bet on it! All the more so that the archetype of the Thinker was becoming simultaneously a miror, a model and a method, if not psychology, semiotics and pragmatics of creation. Not to mention that it lent itself to the supplementary role of a graph.

How to resist... how not to carry through a reflection that would make us plunge into the semantic web. With a condition: our undertaking had to be kept open so it could be qualified everlastingly as creative (if not pataphysical). We therefore had to go from one opening to another. We had to keep being open ourselves to avoid complexifying the creative process with some other existing theories.

We have decided to rely on the wisdom of the model. Therefore it is him or her that will give us lessons of creativeness. And it is their teachings that we will present as a first step. Thereafter we will try to explain how our model projected itself into the semantic web with the help of a new partner: the Machina Mundi team.

The creative process can be seen as a mise en abyme containing three depth levels of the self: the author, the narrator and the actor, three levels that parallels the correspondences already mentionned.
This implies that our graph is generating three fractals of the self, as so many Russian dolls. Fractals that implies that the communication questions and their correspondences can be asked for each level of the self.

Let us test the model with the story as a literary genre.

The self level of the Author reflects on the management of the writing project.

The self level of the Narrator reflects on the management of the medium and the message.

The self level of the Actor reflects on the management of the characters and the events of the story.

Profiles of three literary genres: poetry, narration, argumentation

To insure a full understanding of Cogitae as a fractal phenomenon, in all its simplicity, a phenomenon that resonates from all sides as a constellation would, let us recall that the fractal levels of the self, Author, Narrator, Actor, also correspond to literary genres.

Could we not consider that the Author has poetry as a counterpart, the Narrator has narration and the Actor has argumentation. Thus the assumption that through Cogitae one could make a synthesis of all the literary genres.

Robert and I, we wanted to test this assumption. That is why we used three genres of literary works: a book of poetry, a forestry tale and an essay, that I had authored. What better way than to experiment on one’s own work.

So let us follow our experiment.

These are the books that I used to test the nine questions of the Cogitae’s model.
For each of the works, I answered the nine communication questions acting as an Author. So I stayed at the upper level of the self. Regardless, since for each level the way to proceed stays the same. That said, to facilitate the responses to each of the questions, we blew them away into a multiplicity of possible answers. For each we created a cloud of answers, like an open portrayal of all the possibles.

Answering each question for the three works, as an Author, I selected the appropriate one, thus tracing out the route from the first to the last.

Finally, for each answer of the nine questions about the three works, we did a Google search and gathered a cloud of references and pertinent datas.
But we also wanted to go further and explore more of the potential of the web as a source of information. For that we needed technical help and some kind of automated tool that would not only gather data but also winnow them out to keep only the most pertinent.

**How to reconcile art and technology?**
The Web is like an information ocean littered with gyres of junk sargasso. How can we extract the very substance of it? If not by working from an open, flexible and universal model. Each user will have his or her own answers. Following our example, a user can retrace the same itinerary in customizing his answers to the 9 questions. The results we had with the previous three books would not be the same for another user who would pursue his own objectives. Upstream and downstream, his understanding of each question would be always relative to his own knowledge and to his search of new knowledge. Perusing the cloud, that is cloud computing in search of new knowledge is one way to reconcile art and technology.
I will now hand the floor to Robert for this last part of our presentation.

**The making of our first ontology**
Our first incursion into the world of the semantic web was the making of our first ontology. As you know, in computer science an ontology is a set of concepts organized according to a small number of hierarchical and logical rules representing a precise domain of knowledge. A famous example of that is the pizza ontology that serves as an entry level model to learn to build ontologies.
The problem for us was that this model and other that we examined were based on a set of monosemic lexical concepts that are appropriate to describe a domain of knowledge but are unable to account for the dynamic of a process, in our case the creative writing process. So we had to resort to build our ontology on a set of statements, short sentences, syntagms, arbitrarily chosen to represent the creative process as it was inferred from the clouds of answers mentioned previously.

So here is what our ontology looks like. For each of the nine question-concepts of the Cogitae’s model we have an independent set of hierarchized statements, some general, some very specific, that describe a particular work in a literary genre, that is a childhood story.

**The logicality of our ontology**

Obviously our ontology has not the rigorous set of logical inferences of the pizza ontology but answers rather to a kind of fuzzy logic opened to interpretation. Indeed, creative writing is not based on a finite set of concepts like the pizza world. One cannot describe the creative process with a lego like
model of interlocking blocks. So we thought that the best way to proceed was to rely on intuition and see where it would lead us, the final goal being to verify if our ontology would be functional from a programming point of view.

The WHY question and its embedded folders (French and English version)

![Ontology of the WHY question]

Obviously, there could be many other configurations of answers to the WHY question. I guess any of you could have one as good as ours.

Exploring the web

Thereafter we needed to get some technical help to put our ontology to good use in exploring the web. But first, we had to clarify our objectives and our expectations. Indeed, what did we want to get from the web that the usual browsers would not already give us?

Well, we wanted our queries to be answered more precisely in regards to any knot of our ontology. We wanted our search on a given subject to be more exhaustive and finally we wanted our research to be more expansive, in the broadening sense.

Happily, we happened to meet a group of enthusiastic web experts that were ready to help us to build a browsing tool that would meet our goals and requirements. But first, before I talk of the gathering of data, let us examine briefly how the information is organized in the gazillion containers of the web: pages, databases, big data, that populate the web.
### Information readily visible on a web page

The hyperstructure of a web page conceals an incredible quantity of information for whomever pays attention long enough, information that can be cross-referenced in databases. Here is a table of the most frequent types.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>Semantic indication to be linked with the content of the site</td>
</tr>
<tr>
<td>language/languages</td>
<td>Indicates that it could be possible to exploit semantic terms in other languages</td>
</tr>
<tr>
<td>menu of topics/sections/categories</td>
<td>Keywords that correspond to elements of the site but that are also used by other resources</td>
</tr>
<tr>
<td>contact</td>
<td>Can indicate the author, the origin and geolocalisation</td>
</tr>
<tr>
<td>publishing platform</td>
<td>Standard format that allows the development of exploitation scripts for numerous sites of the same platform</td>
</tr>
<tr>
<td>multimedia</td>
<td>They can be described as radio, television, movie, video, picture, gif, jpeg, png, mp4, etc. attributes that help set up an object database</td>
</tr>
<tr>
<td>agenda</td>
<td>Helps to build a database that corresponds to events that can be broken down into author/subject/location/duration/qualification, etc.</td>
</tr>
<tr>
<td>text accompanying pictures</td>
<td>Textual indication of the picture subject that can be linked to other texts concerning the same subject</td>
</tr>
<tr>
<td>catalog</td>
<td>The catalog is in itself a database of articles, books, journals, texts, pictures, etc. that contains a textual description for each element.</td>
</tr>
<tr>
<td>list</td>
<td>La présence de liste permet d’augmenter la taxonomie originelle en lui additionnant les termes sémantiques trouvés dans les listes et dont on indique la provenance</td>
</tr>
<tr>
<td>software</td>
<td>Softwares can have many attributes and functions: browsers, videogames (character, role playing, translation, etc, that can be linked with our categories, the categories of other sites, some other person, etc.</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>bibliography</td>
<td>If a bibliography is the documentary source of a given topic, it would be useful to mention the subject or the category which describes each work for further use as in a <em>see also</em>, for instance.</td>
</tr>
<tr>
<td>biography</td>
<td>Each biography is an important part of a large amount of historical, conceptual, geographical data, etc.</td>
</tr>
<tr>
<td>course and training</td>
<td>They are a good way to improve the specificity of the terms of our ontologies</td>
</tr>
<tr>
<td>hypertext</td>
<td>dictionaries, encyclopedias</td>
</tr>
</tbody>
</table>

**Information less readily visible on a web page**

With a right-clic of the mouse on a web page, choosing *info on the page* or *inspect* depending on the browser, it is possible to get some metadata about the page: title, description (set of syntagms that could become keywords), keywords, default language, font id, origin (locate), type of content (articles), but also about the content of the page, for instance the pictures.

![Web page metadata](image.png)

With another right-click, choosing *display the source code*, it is possible to get a wealth of supplementary information. This information is generated automatically to some extent by the software and the page making platform, but is also recorded in part by the developers themselves.
We can see that this page has been written in html and contains a list of indications about the assembly of the elements in the hyperstructure: links, pictures, etc. That kind of encoding speaks to the intelligent agents that are programmed to read it. The question is therefore to take advantage of the W3C standard and the interoperability of the various formats to program agents, robots, that will be able to read any web page. And, as you can imagine, this is no small task!

**Strategy to exploit the web resources**

Our strategy consists in developing taxonomies, lexicons and conceptual grammars that constitute the field of knowledge of a creative writing ontology. In assigning to each term an attribute like author, genre, school of thought, style, format, number of pages, etc. we will be able to produce more and more complex cross-references in our database. Of course, provided that the robots encounter syntagms that correspond exactly to the terms we have defined in our taxonomies, their synonyms, their translation in other languages or their equivalents, as they were encoded in our system.

**Qualification of the primary resources**

Robots execute the operations they have been programmed for. One of these operations can consist to browse the links of a page in a recursive or cross-recursive mode, indefinitely, so that attributes can be associated to the new pages inherited from the primary resource. We have then an assembly of
database items that have common attributes. Invoking an attribute by means of a results display tool will show the primary resource containing the attribute but also all of its children.

Performing queries by means of our terms
We can rapidly increase the number of resources contained in our system by using our terms in library browsers, bookstore catalogs, academic databases or even in pertinent sites or blogs. These resources can be garnered by browsing the source code of the pages and linking them to our subjects, attributes and conceptual grammar by means of definitions and actions scripted in the robots.

There is no limit to creativity in the exploitation method
Many methods of data collection are conceivable, the available technologies exceeding largely the possible exploitations by human beings, whether they are specialists or not in a particular field. The aggregation of content in real time must be controlled. It is impossible for a human operator to carry out the various tasks of data collecting around the clock, that is why we need a user-friendly interface to feed and drive robots. Several of these interfaces commonly called content management system (CMS) already exist, like Joomla!, Drupal, DotClear, Wordpress, etc. Each one is suitable for one or many types of activities and each one has also constraints that limit the scope of project that exploits databases like ours. With the help of the Machina Mundi team, our friendly web specialists, we are contributing to the development of our own CMS to increase the efficiency and the performance of the data gathering. This is made possible by eliminating useless or superfluous functions, making the system lighter, enabling us to develop numerous customized display forms to feed the robots.
At first, these forms will be filled partly by human operator, but eventually fields corresponding exactly to certain metadata will be filled automatically by machine operators.

How it works concretely

Illustration 1: Information page
The first tab of this form indicates that this URL is a page of a parent site (15 sites that are indispensable for writing on the web and in blogs, if I may freely translate the title). Thus with this form we get the title, the URL, the parent, a description, a brief overview, the date, the importance (quality of the resource) and the device name.

These information are set in stone for the robots. They will go back to this page systematically, browsing the page, weighing the text. And this page describes the link toward 14 other pages.

Illustration 2: Resource qualification tab

Already we discover an author whose Wikipedia article we discovered via a link on this page, a link that also leads us to his personal page, to other works, etc. The sites listed on this page refer explicitly to syntagms of the Cogitae’s ontology which are thus selected. This page becomes an associated resource to the Écrire pour tout le monde (write for everyone) entry as many others will, as human or machine operators enter them in the system.
The robot waits no time to spot whatever constitutes a link. It did not find keywords, documents or person names in this precise page, but it comes back with all the links, the pages, the queries, the rss feeds and the pictures contained in the page. Each of these items can become a new form which will be analyzed in turn and which will be linked to its original resource by the sharing of qualities, attributes, syntagms, paradigms on which the Cogitae’s system is based.
Illustration 4: Information on the author of page 15 (in Presse-Citron)
The author of the whole Presse-Citron site (which can be translated by Lemon Squeezer), Éric Dupin, is actually Éric Dupin de Beyssat. A robot has rapidly found his Wikipédia page and extracted a wealth of information with the Wikipedia standard, like the date with the current format (26 mai 1955) that was transformed in the ISO format (1955-05-26). This person is therefore linked to the page: 15 sites incontournables sur l’écriture… (15 sites that are indispensable for writing on the web and in blogs) that acquires to some extent a connection with the sites listed in the next tab.

Illustration 5: Onglet Qualification de la Personne

The robot has found the sites on which Éric Dupin de Beyssac has already published some content. If they were not already there, it would have added them to the system. At a higher level, an operator qualifies the works of this person, works that are thus associated to some other works with the same type or genre of writing. Éric Dupin will then be associated to some other content providers, in blogs, doctoral thesis and more, if the topics of his works correspond to the topics of their works. One can then assign a school of thought or a literary movement to the persons that share the same attributes and topics.

That is how the system adds successively layer after layer of knowledge, horizontally and vertically, knowledge that can be offered to and exploited by a specific user that can select its preferences through various forms.
**In short**

The creation of this tool can be likened to creative writing. We have to get benchmarks and get the job done. The idea is not to tell but to program a story by means of programming languages, alphabets, syntagms, syntaxes, styles and any other feature contained in the character strings. In this project, it is necessary to list the pages of the selected sites as each page develops a certain aspect of the proposed thematic. Likewise the links found in a site tend to refer to the thematic of the said site; they increase its content and its relevance. In perusing the source code of the pages, the robots gather rss or xml extensions (feeds), the results of queries, documents (doc., pdf., odt., xls., etc), pictures (jpg., png., gif., bmp., etc.), tags and the persons mentioned (dc:author, <name>, <family name>, etc. A robot can then browse all the references found and create as many new resources as are related to the primary resource.

Let us emphasize that it is also possible to create new connections in other languages using the Traductions (translation) tab. If the site already includes version in other languages, the address will be found under the Alternatives tab.

**Now, how do we plan to exploit that database**

The ultimate goal of this project is to transmit knowledge without a person itself having to browse the web indefinitely. Therefore we will use the numerous available technologies to build a dynamic web site (in which pages are created by clicking) that meets the presentation structure of the data. Specialized programs will extract the data from the base according to the date they were created, the author, the subject, the format, the type, the language, the location, the relation with other subjects, all elements that will have been defined in our ontology.

**Expected results**

For example, that search platform should enable us to find several types of information that concern an author and his or her work.

- All the authors that are interested by a subject;
- All the subjects treated by an author;
- All the authors that have adopted a given genre;
- All the genres explored by an author and all the documents related: web site, interviews, publications, critiques, e books, press article, articles of journals, photos, etc.;
But also information on the writing process and the content of literary works.
Source of inspiration;
Narrative technique and patterns;
Sentence construction;
Style analysis;
Examples of a given topic in different works: childhood memory, extraordinary situation, final outcome;
All the references to a location;
All the references to a given period;
Etc.

So that will be all for our presentation. Thank you all.

Afterword
If you do not mind, let me add an afterword.
We started on this work about one year ago and I must admit that, at the time, we did not have the slightest idea of what an ontology was in the IT sense, neither did several of our friends that worked in the field of computer science. We were attracted to by the semantic part of the expression semantic web. It took some doing before comprehending that the semantic web and its ontologies was a way to logically organize knowledge that makes knowledge more accessible, more rapidly to more people. Actually, I believe the semantic part is still in the hands, if I may say, of humans.