MATRIZes

The Cybercultural Scene in Contemporary Journalism: Semantic Web, Algorithms, Applications and Curation

Elizabeth Saad Corrêa[1]
Daniela Bertocchi[2]

Abstract
This paper discusses the different cybercultural aspects within contemporary journalism experiences, mainly from the emergence of the Semantic Web and the technical functionalities that result from it, such as the use of algorithms and applications. We explain the basic concepts of this technical set and it's correlation to the Communication and Journalism fields particularly. The possibilities of approximation of this scene within journalistic activity are presented due to examples which are already in application with international journalistic markers. We come up with a configuration of the curation role for the professional that will act within this new cybercultural context.
Keywords: Cyberculture, Digital Journalism, Semantic Web, Algorithms, Curation

Introduction

We aim to think communication happens when machines also communicate, especially in the journalistic information context. This concern rises as a result of the recent and increasingly present debates about the development of the Semantic Web (SW), one of Tim Berners-Lee (Bourenane, Szoniecky, &Saleh, 2009) most well-known predictions which now begins to reach it's stage of consolidation (Adolphs, Cheng, Klüwer,Uszkoreit, & Xu, 2010). From the SW, we are also interested in discussing the debates about the use of Algorithms and Applications (apps) in the journalistic process, exploration which discloses an irreversible trend in the main highlighted essays of the web as (Rabaino, 2011 and Carr, 2011).

The SW, apps, algorithms, databases, among other things, are contemporary cybercultural manifestations and change the ways of sociability, therefore, we are faced with another discussion and reflection field, not to say pressing changes. It is in these terms that we will develp our argument.

The first part of this paper explains the Semantic Web status as a wide process. By defining it, we privilege the point of view of communication, language and cognition over the technical view concerning the subject. We believe that starting this paper presenting the SW in this framework is necessary for at least two reasons. The

1 Professor at University of Sao Paulo and PPGCom-ECA. Coordinator of both Digital Communication Graduate Course (DIGICORP) and COM+ Research Group.

2 Daniela Bertocchi is a Ph.D. Researcher in Communication Sciences at University of Sao Paulo, Brazil. Master of Cyberjournalism from University of Minho, Portugal (2007).Researcher of the group COM+.

first one: the specific literature production about the SW has been prolific in the fields of Information and Computer Sciences, and also in the Engineering. It has been usually prolific in the English Language; however, it is scarce in the Communication field. Therefore, we regard it as essential for discussion in the debates concerning Digital Communication and Cyberculture. The second reason comes as a result of the first one: the concept of SW is still a subject in it's infancy among searchers and communication professionals, which helps us firstly explain it's operation logic before making a profound study of its specifications. In this first part, we also correlate the SW with Algorithms and apps among other contemporary ways of Communication.

In the second part of this paper, we discuss the scene in which the machine includes data and, one step further, correlates and interpretes it showing specifically how such scenes set in journalistic environments.

In the third and last part we reflect on the journalist's role as a communicator in their own field of Communication in this context, which makes clear the excess of available data on the net and the several attempts to organize and give meaning to it.

The SW subject has been discussed in different ways among the academics of Communication and Information Sciences and it has been appropriated by countless Media Representatives as "the future" of the World Wide Web. While carrying a short and unsystematic bibliographic search for the term "Semantic Web", in academics as well as in Google itself, we find out a myriad of expressions which translate, in our opinion, mistaken views concerning their social impacts: the idea that the web will have autonomous "intelligence"; that it will be a repository of the world's knowledge; that it will be able to recognize the individual patterns of use and personalize contents for each user in order to increase this universal intelligence; a system of digital marketing leverage; an extension of the user's memory and, at the same time, an individualized file of its knowledge, among other things.

In general, we see the mystification of the SW, it is erroneously linked to just an idea of the "Web 3.0", announced as the brand new stage for the World Wide Web, it is evident that the SW will promote significant changes in the use of the web – it has already been happening in many cases – but we intend to reflect here on such transformations as a result of a process which has it's setting installed in the back office

of the net (Algorithms, Softwares, Database Systems, Applications, Machine Language, etc.), a mechanic process that will have a socially transformative dimension through the practice of the communicative process. In other words, through the performance of a communication professional as a representative, a nudger, a mediator, articulator and a human behavioral analyst at this organization stage, who adds organisation to the huge information mass which circulates in the Cyberspace. We call this person an Information Curator someone who puts perspective on the data.

In fact, when the information overload becomes unbearable, the user calls upon experts by delegating them the task to organize and give meaning to the data. When it comes to news information, it is expected that the journalist carries out this task. For some people, in the Web 3.0 scene, the Algorithms will be the great experts capable of accomplishing this mission and will even replace the human editorialization. According to our view, a socially relevant SW will not speak for itself, but through Communication. That is to say, without a communicator as a representative of this net beside the machine, there is just a Totalitarian Semantic Web. So, it is important to rethink the journalist and informative company position according to this next perspective.

Part 1 – What are and how do the Semantic Web and Curation Algorithms work?

One of the most accurate definitions about Semantic Web is the "unambiguous web"[3] (Siegel, 2010). This vision places the SW as a digital "unambiguous" space where machines are able to understand the meaning of data in different contexts and without any overlapping interpretations[4].

It is understood that, for example, the Brazilian Soccer Team *Esporte Clube Vitória* has it's headquarter in Salvador and not in the city Vitória da Conquista, in Bahia, a place which is far from the capital Vitória in Espírito Santo State. It is a well-known fact that there are plenty of girls who are called Vitória, however, just one grew

3 It is important to remember that Communication works, on its different narrative processes, by the use of ambiguity to persuade, make others believe, inform, contextualize. Hence the importance of the communicator as a mediator and active pusher of the SW applicability.

4 The definition which is most quoted is the following: "The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (Berners-Lee, Hendler & Lassila, 2011, p. 35). Breslin, Passant & Decker (2009) define it as "The Semantic Web is a useful platform for linking and for performing operations on diverse person- and object-related data gathered from heterogeneous social networking

up and reigned for over sixty years in England. There is, in Foreign Literature, a work called *El Pensamiento* by Francisco de Vitoria and another, in Brazilian Literature, about the *Vitória-Régia* Legend, which is in a Biological context the *Victoria Amazonica,* a flower that has a little or nothing to do with the *Esporte Clube Vitória.*

To humans, such subtleties about the term "*vitória*" are easily seized as the context and conditions of enunciation. To the artificial machine intelligence, this kind of apprehension is harder. Hence the need for a -- Computational, Linguistic, Communicational – project called "Semantic Web"[5].

The most obvious utility of a SW is reveiled when we try to recover some information in Cyberspace. Nowadays, such a process is linked to Search Engines based on Algorithms like Google. When asking the machine: *Who wrote "Principia Mathematica"?,* it recovers two kinds of information: the three volumes of *Principia Mathematica* by Alfred North Whitehead and Bertrand Russell and also *Philosophiæ Naturalis Principia Mathematica* by Isaac Newton[6].

Therefore, to a Computational Search Engine, both works that gave rise to Modern Mathematics and those which presented the main Theories of Physics are, roughly, the same thing, though a greater knowledge states that it could not be considered true. It is missing, then, a more sensitive view. Or, preferably, it failed to explain to the machine that the same term can have a variety of meanings and then failed to indicate those meanings in each type of context. It is lacking Semantics, or like some programmers would say, there is an absence of Ontology.

According to the conceptualization of the World Wide Web Consortium (W3C)[7], Ontology

> (...) defines the terms used to describe and represent an area of knowledge. The Semantic Web needs ontologies in different levels of structure. These serve to specify descriptions for the following types of concepts: classes (usually things) in various fields of interest; the relations that can exist among things; and the properties (or attributes) that such things may have.

King & Reinhold simplify this by saying that ontology is a "collection of concepts,

---

sites." Read Bertoocchi (2010) in order to get other definitions.

5 It is worth highlighting the idea of a "project", in other words, the SW does not occur as a quantum leap, as many would preconceive, like suddenly getting through to a new stage on the net. Actually, the SW is the Web itself within another setting.

6 This search (Who wrote Principia Mathematica?) was accomplished through Google March 15, 2010.

7 Global entity, based in The U.S., which defines the usage patterns and protocols of the WWW and that, nowadays, has a specific group to define SW patterns.

organized in a hierarchy of categories, combined with relations among the concepts, in order to reflect the vocabulary of a knowledge field" (2008, p.8, our translation).

Why does the Semantic Web need ontologies? Would not a dictionary or a well-structured taxonomy be enough to explain to the machine the nuances among the terms? The dictionary has the purpose to define concepts. The format allows it to present more than one definition per term. There are dictionaries about synonyms, antonyms, and specific terms in different areas. The Thesaurus works on the logic of synonyms by approaching similar concepts. The taxomony arranges concepts into hierarchy. However, it is the ontology that creates semantic relations among terms and, certainly, with ontologies; an unknown number of them can be created like a mind map.

A first issue emerges here when focusing on the digital journalism field: the need of a specific construction of ontologies for the field or a lack of it. This issue will be discussed later. The ontologies can be enriched through folksonomy (social tagging) (Angeletou, Sabou, Specia, & Motta, 2007). The free tagging of contents is dynamically updated by a pleiad of users, what makes it rich in jargon, popular expressions and ordinary terms that can be included in formal and previously structured ontologies which acquire characteristics of the natural language: the multilingualism, ambiguity and inventiveness.

In a web which exempts ontology, the user navigation is usually conducted by the Search Engine which is based on an Algorithm (a standard scanning protocol of the avaiable data content in the net). If we consider the omnipresent Google, the spider scans - the tagged and organized data in the metadata space provided by the structure of the HTML language; it does the same successively to all the addressed pages in the net; it organizes these keywords in an index data. When a user feeds the tool with a searching term, the system processes such keywords lists and tracks the websites which contain the same list. The result is presented through an Algorythm-the PageRank, which ranks them by how often the word appears on the page, by the time the page is online and by the number of other pages that refers links to the same results.

Therefore, the communicator role in this process depends on it's domain

constructing a page content open to tagging within programming rules and it's competence on making this content visible to the searcher by using SEO techniques (Search Engine Optimization). It is a process that the communicative action is determined by the search engine and not by the meanings assigned by the public to the product, neither is it determined by a service of a brand or company and the relation among them and their public.

It is according to this view that folksonomy and ontologies consist of key elements to introduce the meaning of the real world into the searching process and the net use. The basic approaching elements between the data generator-user are through a changing process of communicational meanings, semantically included in the process of systems searching. The presence of a communicator in the construction of folksonomy and ontologies, and in the instruction of the constructive process of pages with semantic links, will be fundamental to the SW consolidation.

The searcher Silvia Laurentiz (2010) believes that the biggest SW challenge concerns the issue of the creation of ontologies. "Defining ontologies", she writes, "provides a set of automatic or semi-automatic methods and techniques for knowledge acquisition by the use of texts, structured and semi-structured data, relational schemas and other knowledge bases" (Laurentiz, 2010, p.17, our translation).

When applying Laurentiz' words to an ordinary scene of journalistic writing, we can see the idea of a universal prior structuring, the onotology, becomes displaced from the journalistic process that requires both newsworthy editorial criterias and those relating to SEO construction of titles, menus, sections, tags and metadata. It would be a scenario of oppositions between automatism and journalistic praxis.

We must reflect here what the communicator, journalist and journalistic business role is to keep such sets of values in a process which aims towards the universalization of them and that is constructed, due to the lack of options, based on a set of subjectivity (those which are the ontology constructors). We have observed in the different writing práxis, with a significant presence in the net, the option for a kind of "middle path" to enter the SW world without abdicating the specificity of the field. This path does not prioritze the ontology based on a semantic performance of

the journalistic marks in the net, but emphasizes the owner construction of CMS's (content management systems) - the publisher systems. We are discussing, therefore, the use of "semantic markers" (not ontology anymore) incorporated into the CMS itself, in other words, a set of terms and words, strategically defined according to the business model and the editorial line by the informative company owner and user of the CMS, which work as parameters to any journalist during news production. Those parameters consider the characteristics of the publication audience/segment and the SEO indicators. The CMS, by default, leads the journalist to categorize the content produced within this set of markers.

These practices make the association of the Semantic CMS possible in conjunction with other emergent practices in the 3.0 journalistic scene which includes the Algorithms and the apps as described below.

To conclude, by CMS, ontologies, algorythms or apps, the raw material of the entire journalistic-communicative process keeps focused on the word as world significant and news values. It is through the word that ontologies are constructed and the searcher action is done. The journalist marks the content and the Algorythm customizes (by hashtags #) the most informative offerings. It consolidates the role of the word's artificer.

As a parallel to this, on the web or through applications for tablets and other mobile devices, there is a significant presence of algorithmic solutions to deal with the informative profusion. The searcher Google (and other more segmented searchers)[8] and the Social Network Facebook are based on curator-algorithms[9] that decide which information will be available. The aggregation systems, based on the organization of information flows in RSS[10], use the simplified algorithm of which GoogleReader is the best example.

Recommendation Systems like the one on the virtual bookstore Amazon, or the ones from music websites like LastFm and Spotify, also use fairly simple algorithms. There are algorithms created to support business models and that became specific

[8] The Addict-o-matic which favors the search of the buzz on social networks, the DuckDuckGo that searches contents in websites like crowdsourcing and the Icerocket that searches at real time on the net are also included in this category.

[9] We took the term "curator-algorithm" to identify this kind of automatized action.

products based on personal feeds and/or tags defined by users: as was the case with Zite and Flipboard, costumizable magazines, unique for each user and yet exclusive to tablets; the Paper.Li, which organizes personal newspapers from users browsing social networks; the Scoop.it which creates a dynamic content from user defined-themes; the Storify that constructs "social narratives" and makes the result go viral; the Social Pulse, of the news agency Reuters, which presents the most tweeted and retweeted news by a network of influencers; the Washington Post Social Reader, a news aggregator which is also customizable; and the tools used to organize information through images, bringing the innovation of social learning into the construction of visual narratives, such as with Pinterest and WeHeartIt.

We mentioned here only the most popular examples of what we call organization tools of informative abundance by algorithms. What we must reflect on in lieu of this availability is its strong curatorial role, already intrinsic to the tools, and the idea that every algorithm is the result of a human process with choice criterias predetermed based on a context of information offered, as is described below.

The term "algorithm" is a Latinized word which comes from the name Al Khowarizmi, an Arabic Mathematician from the 19th century. In computing, an algorithm is "a procedure designed to accomplish a specific task" (Skiena, 2008, p.3, our translation).

It is accomplished through a computational step-by-step, a programming code, performed on a given frequency and with a determined effort. The concept of algorithms allows us to think of it as a precedent which can be performed not only by machines, but by men, expanding its acuity potential associated with customization.

Therefore, designing an algorithm is creating a number of instructions with the purpose of solving a problem. In the Facebook Timeline case, the algorithm is especially used to instruct elements (data about other users) by importance criteria (defined by the programmer of the company). According to Google, the algorithm tries to solve the data comprehension problem and return related occurrences. For Zite or Flipboard, the goal is to make a digital magazine with relevant contents relevant to that user, and so on. In the digital communication scene, strictly

[10] Really Simply Syndication

speaking, the algorithm works with the mission of purging unnecessary information, offering only what the user would judge as more relevant. It all acts in accordance with a defined business model or access to information previously determined by the owner of the algorithm.

When considering the use of algorithms to decide criterias of newsworthiness in journalism, like the curator-algorithm of Storify, we would be endorsing a process which dispenses the "journalistic" variables and relevant social event typical of the field.

We can deduce from an operation with this kind of curator-algorithm that the journalistic activity in this context is reduced to the production of information and it's availability in the huge web information bowl. Will just the machine and the intensity of use of those platforms by each user be responsible for the decisions made about relevance, editorial topic and journalistic Mark that will be offered to the reader?

Part 2. The machine which comprehends and the editorial department that deals with intelligent machines

We have been facing a Cybercultural Scene of the informative world which provides, for the comtemporary editorial department, a World Wide Web full of information never seen before; a medium-term proposition of another net – SW; an imminent reality of a construction process of the public opinion ruled by platforms which are based on curator-algorithms; and a possibility (characterized as an investment) of an intermediate process with Semantic Marking of the CMS's and the Journalistic Folksonomy.This scene is, at the same time, expected by different analysts (as described below) to be the Future of Journalism and also begin a new trend of Paradigmatic Changes.

The information overload, which the World Wide Web fosters, has been the *leit motif* to justify the emergency of the Automated Systems due to the modeling process with a clear human intervention. Here are some copyrighted thoughts on the subject.

Informative overload, Information Anxiety and Information Diet (Turkle, 2011, p.242; Pariser, 2011; Johnson, 2012, our translation) are expressions which began to

emerge in the 1970's[11], since then, they have started to be included in the Communication Searchers' Discourse and even more in Non-Academic and Media talks. It is the result of a moment when Scientific Academic Institutions, School, Libraries, Museums, Organizations, all kind of corporations – not only the traditional means of communication – share digital data in the net and multiply the information supply for society.

It is known that there is an information abundance and a spread of multiple narratives about world events created by different sources and presented in formats to suit all tastes. Nevertheless, there is no guarantee of *data* transformation into *information*, nor into *knowledge* on the part of man[12]. The Media Specialist Neal Gabler (2011) says that society lives in the post-idea age, in other words, the individuals became big on facts and information collectors, however, they cannot develop a critical and deep thinking about said facts. The Media Commentator states that the Digital Age delivered us to "well-informed ignorance".

Weinberger (2012) identifies in this information overload scene a "knowledge crisis". According to the author, in the pre-internet world, information remained clearly and materially localized (in books, libraries, newspapers) and allowed the construction of knowledge through the "hard work" of scholars who became, consequently, experts in certain subjects. In the digital context, according to the author, we have been facing the opposite: information is spread in a disorderly fashion; it is provided by amateurs, plagiarists and users who consider a good content the one which has the biggest number of thumbs indicating "like". Even so, he reiterates, the Means of Communication, Companies, Governments and Science have benefited from the available data in the net, principally, on account of readers contributions, who make the data easier to find and understand to many other readers while also contributing to the process. (2012, loc. 126-132).

In other words, the knowledge crisis rises at an age of knowledge exaltation. The networked knowledge, though less accurate, becomes "more human".

[11] Alvin Toffler introduced the idea of information overload in 1970 through his book *"Future Shock"*.

[12] We mentioned the famous DIKW pyramid (data, information, knowledge and wisdom) created by Russell Ackoff in 1988.

Analyzing the same scene through Social Philosophy, the thinker Pierre Lévy who defends that - since 1991 when *L'idéographie Dynamique* was published – the programming approaches the human thought operation. He always thought that Computational Programming lacked the development of the representative and expressive side of language (the symbolic, proper of man writings). Lévy suggests nurturing the familiarity among intelligent machines, information builders and society as a whole, and since 2006 started to make a profound study of what He calls "*la langue de l'intelligence collective*", entitled IEML (Information Economy Meta Language).

The IEML is an artificial language of the collective intelligence which is a) manipulable by computers and b) capable of expressing semantic and pragmatic nuances of natural languages. It is the SW being understood not only by a computational view, but it also fits within a Philosophical and Social Perspective. Or, like the author states, it is about a work, the "*human-centric social computing*" (Lévy, 2009, p.32).

According to Lévy's proposal, the Semantic Marking and Journalistic Folksonomy, which have recently begun to be practiced inside big editorial departments, should be enlarged to the web as a whole; consequently, a "Semantic Space" of social construction would be created.

In the same Philosophical-Social context, there is the researcher João Fernandes Teixeira (2010) who goes further and talks about the post-evolution world - or post-human, whose main characteristic is the impossibility of the human body and brain to keep developing and leading the species to resort to parabiosis (the association of men with forms of dry life):

> "Until the arrivel of the post-evolutive worlds, we have gotten into a situation where machines bring us a lot of unhappiness, because we are just enlargements of them and we cannot parasitize them. However, it does not mean the post-evolutive world's arrival will, necessarily, make us happy. The parabiosis is not the achievement of a Utopia" (Teixeira, 2010, p.12, our translation)

In a world of digital convergence, supermachines or super cyborg figures with enormous computing power would process a tremendous amount of data. These supermachines or hybrid machines "*will have a brain more powerful than the human*

one, so they will be able to process data and statistical programs which our present supercomputers yet cannot do." (Teixeira, 2010, p.73-74, our translation).

In other words, machines will not only process huge data bases and information but they will also produce *knowledge.* To Teixeira, this will be an *artificial knowledge* built through the correlation of databases -- including random correlations -- that will generate forecasts and projections in different fields. "*A way of transforming information into knowledge is by crossing data*", he states (2010, p.75. our translation).

We enter into a world where the machine not only comprehends data – which is the main goal of the Semantic Web – but also correlates it. This will generate new propositions to humanity according to the author.

Between the informative overload and the emergency of a society informed by machines that are intelligent and established as a human mind extension, we can see the Journalistic Field - even being a conservator - does not place itself outside the scenario. On the contrary, some of the most important global journalistic brands - *The New York Times*, *The Guardian*, *BBC, Reuters* among others, search for approaches to this Semantic World which is pulverized by apps and platforms.

At the end of 2011, Jeff Sonderman (2011) who is a *Poynter Institute* analyst and known in this field for the seriousness of his studies published three tendencies that would guide Journalism in 2012. They all reflect in some way the use of Algorithms and Semantic Systems. They are:

- the emergency of narratives beyond authorial texts, supported by a variety of informative layers such as Storify[13] and ProPublica[14]. Therefore, the tendency is to make the audience become a contributor to obtaining news, the journalist assumes the role of a society listener and curator and the news is reported by the society through a journalist and not as the traditional narrative from a journalist to the society;
- the use of Facebook and other platforms of Social Media (typically consisting of algorithms) as the best means of publication and spreading of news; and
- the majority of tablets and e-readers (representing the mobility of the reader) as the main access to informative means.

[13] www.storify.com
[14] www.propublica.org

The Columbia Searcher C. W. Anderson states that Journalistic Curation performed by the Algorithm can be inserted into an emergent study field – the Sociology of Algorithms (Anderson, 2011, p.529, our translation). When talking about "Journalism of Algorithms", Anderson says the Algorithm is performing a sociotechnical role increasingly important in mediation among journalists, audiences and media products. This mediation has both normative and sociological implications. His research was extensive in the Editorial Departments of the main North-American Newspapers and concludes with a positive view of the use of Algorithms: it is an effective techno-social way to make the newspapers know and understand their public better by including it's use in the DNA of the Contemporary Journalistic Work.

The MediaBistro consultancy, through it's blog, *10.000 Words – Where Journalism Meets Technology* (Rabaino, 2011), also indicated paradigmatic changes in the future through:

- automatization of information collection and publication systems (managed by Algorithms) by reducing the amount of paper in the editorial department and expanding the possibilities of connection and synchronization of content. The use of GoogleDocs, IPads and Evernote is highlighted;
- public sharing of APIs[15] and source codes of informative contents in order to expand the reach of information produced by the editorial department regardless of the type of branded content.
- consolidation of the *open news* capture system, such as the newly formed program of The Guardian, where the summary of the news is posted open in the net to receive the cooperation of readers;
- consolidation of intelligent publishing systems (CMS) based more on Ontologies and Folksonomy than Taxonomies.

Nicholas Carr (2011), an expert author of the technological area, preludes that soon the informative company will be "appificated", to put it differently, directing it's content more and more to accessible apps through mobile devices that add different news sources, reducing the webpages weight of each informative mark. Carr says:

> Today, as a result of cloud computing and other advances, applications look more and more like media products. They are add-supported, subscribed to, continually updated and the content they incorporate is often as important as the functions they provide. As traditional media companies have moved to distribute their wares in digital forms – as code, in other words – they have come

[15]Application Programming Interface is a set of routines and standards established by a software to use it's features by apps that do not intend to get involved in details of the software implementation, but

to resemble sofware companies. They provide not only digital content, but an array of online tools and functions that allow customers to view, manipulate and add to the content in myriad ways.

We can presume, based on these examples that the "middle path" seems like the option for use by some informative companies. Even so, it is evident those still transit into uncertainty when they need to put important decisions in a strategic context and business model like investing in a CMS owner - which implies a cultural and professional change; a content "appification" - causing the branded content denial; or even disinvestiment in their webpage.

Once the scenario has been discussed, we dedicate ourselves in the third part of this paper to ponder the possibilities of a journalistic position for informative companies in the presence of the current Cybercultural Panorama, without questioning the well-known art of journalism surviving. We do propose a resilient and adaptive position in the future.

Part 3. Journalists and their *modus operandi* in a social-mechanical context: the curation

Within objective limits we can assert that, nowadays, a journalist transits in the midst of information overload, the need of reporting the world to society, technological learning, the predominancy of moblie devices as (non) localization of their audience and the proximity, sometimes intrusive, of their old readers. That is, a typical panorama of the current Cyberculture.

Information overload needs organization and contextualization, technological learning requires corporative investiment and intellectual availability, mobility and proximity requires new narrative formats. How much do these requirements approach the current informative process in informative companies, which are based on a classical proposal?

Even when we are able to classify and indicate some experiments carried out by the media or platforms cured by algorithms[16], it seems that in a context of informative abundance in which a machine processes information faster and better, (even being capable of making correlations and in the future providing artificial knowledge) Communication, Journalism and their professionals still remain indispensable for, at least, two reasons.

---

only use it's services.

16 The Publisher Editora Abril, in Brazil, has been developing its CMS owner and brand integrator in 2012 – the Alexandria, designed within the proposed use of semantic markers and journalistic folksonomy.

The first reason: there is no robot that has passed through the Turing Test[17], in other words, they cannot yet "talk" to a human, interacting in a natural way and establishing empathy (or antipathy) as another human could do, either making profound or unpredictable correlations.

It was exactly what the North-American Newspaper *The New York Times* noticed in May, 2011. The newspaper exchanged their robots for real journalists. The newspaper left the automatic content publication at its Twitter (@nytimes) profile, replacing robots by human beings for one week, in order to experience and analyze if an interative human approach would be more efficient to communication, engagement and audience[18]. One of the Editors, Liz Heron, stated that it was all about an attempt to observe how a "cyborg" approach would deal with a human approach.

After a week, at the end of the experiment[19], the newspaper returned to it's old format and robots resumed feeding the Twitter account. Talks with readers about the news had to be stopped ("Reply" abandon) and hashtags (#) removed from the setting. The retweets also could not be made anymore, because the "recommend" decision of a content can only be made by an editor. The news headlines turned out to be the same as the ones published on the website without specific edition to a reduced space that holds only 140 characters. Even without official publicizing, it follows that systems, robots and algorithms cannot dialogue with the audience and have little influence on collective opinion formation despite optimizing the costs and resources of the editorial department.

On the other hand, *The New York Times* itself records a significant success with the postings on it's Facebook page, whose technology does not favor the automation of content. The newspaper itself declares the relationship established with it's audience through comments in the postings is something valuable for it's content and consolidation.

The second reason is: in Semantic Space, intelligent robots notice user behavior and provide only the information which the user itself wants and chooses. It comes with the maximum level of customization, which brings some undesirable implications beyond the clear advantages.

The main point is that the citizen gets into trouble in what Eli Pariser (2011) called as "*the filter bubbles*". He states:

> I am progressive politically, but (...) I like hearing what they (conservatives) are thinking about (...) And so I was kind of surprised when I noticed one day that the conservatives had disappeared from my Facebook feed (timeline). And what it turned out was going on was that: Facebook was looking at links which I clicked on, and it was noticing that, actually, I was clicking more on my liberal friends' links than on my conservative friends' links. And without consulting me about it, it had edited them out. They disappeared. (...) (Pariser, 2011)[20]

Pariser (2011) says that not only Facebook but also Yahoo News, Huffington

17 A proposed test by Alan Turing on a publication in 1950 called "Computing Machinery and Intelligence" whose goal was to determine if machines can think. In the test, there is a man (A), a woman (B) and a judge. The judge will not know that the man is A and the woman B, but will try to identify each one of them through some questions. Both A and B can lie to make things harder. At a certain moment, a computer will replace A or B. If the judge does not notice it, that computer will have passed the test. (Accessed at ttp://pt.wikipedia.org/wiki/Teste_de_Turing)

18 Poynter. (2011)."Why The New York Times replaced its Twitter 'cyborg' with people this week". Available at: http://www.poynter.org/latest-news/media-lab/social-media/133431/new-york-times-tries-human-powered-tweeting-to-see-if-users-value-the-interaction/. Accessed: May,2011.

19 The newspaper did not disclose the results of the experiment.

20 Available at http://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles.html Accessed: June, 2011.

Post, The Washington Post, The New York Times are flirting with personalization in various ways. It means we have been moving toward a world in which the Internet is showing us what it thinks we want to see, but not necessarily what we need to see. "*Your filter bubble is your own personal, unique universe of information".*

The author says we are seeing more a passing of the torch from human gatekeepers to algorithmic ones. And a big problem is that the algorithms do not yet have the kind of embedded ethics the editors did. The algorithms will decide on what we will watch, read, listen to and know, according to what we judge as relevant but not necessarily consistent with what is relevant to society's point of view. Pariser (2011) says that *"we need to make sure they also show us things that are uncomfortable"*. This is the point where we start pondering the formation of a journalist-curator[21]: the establishment of a relationship between sophistication of the algorithm and congruent specialized human intervention (the journalist) in the process of it's construction. The more circumstantial, social and behavioral information become indispensable to the algorithm modelling, the more human participation should be required as a model feeder and, specifically, as a refiner over the lifetime of the algorithm. We observe it, principally, when the curation performed by the least complex algorithms tends to glance backwards: considers the past behavior of the user, what he or she commented, recommended, liked and read.

By being familiar with patterns and preferences, this mathematical step by step brings more similar information related to its user, through a quick and efficient scanning of the databases. The mere human curation (without mathematical preceeding), on the other hand, is freer to look to the future. A content curator is able to include new and unusual perspectives to information, offering to it's users the surprise, the unexpected or simply something the user would never imagine existed in the world and about the world, widening it's own worldview.

The journalistic curation, which uses sophisticated algorithms associated with the expertise of re-mediation of the journalist, seem like being the field for understanding and characterization of the information curation in Communication. The journalist Jane Wakefield reaffirms in an article[22] published by the BBC News: "*Algorithms may be cleverer than humans but they don't necessarely have our sense of perspective*". Anderson (op. Cit 2011) qualifies this process as an "algorithmic promiscuity": the machine's capacity to include and treat computational and human data in the same way, reducing and outlining the latter. Readers of information like that are reduced to their preferences, to what they consume. The agenda setting becomes personal, unique, personalized and determined by their desires. It is an undesired process from the social communication point of view as a whole, since different points of view, sources, perspectives and snips are fundamental to the widening of these readers' worldview and construction of the knowledge of humanity. (Pariser, 2011;

21 Curation has multiple definitions, originally modeled from Arts and Law. We assume the Steven Rosenbaum one in this paper: "Curation is about, he explains, adding value from humans who add their qualitative judgment to whatever is being gathered and organized (even if at first it has been gathered by machines)". According to him, the algorithm cannot lead. What can exist is a work made between machines and people: while computers treat huge quantities of data, humans perform the editorial treatment. (Rosenbaum, 2011, location 395).

22 When algorithms control the world, BBC News, Available at http://www.bbc.co.uk/news/technology-14306146 Accessed: 14 February 2012.

Basulto, 2012).

By reflecting on journalistic curation we assume that the professional, due to their experiences with the ones who want to and have to be informed, has means to important variables about procedures of choosing, tastes, opinion formation direction, social patterns changes and other data, which make the use of algorithmic models as little as alienating as possible. Consequently, they become more appropriate to the social function of the Communication field. It seems that the scenario, initially different, is propitious to one more paradigmatic jump in the journalistic field and to the evidence of a specialized human role as a necessary bridge between information and society.

When thinking about the dimensions of action of the journalistic curation, it is possible to deduce that despite the strong possibility of curation performance by any connected citizen, in the Digital Communication field this potential is reduced to a group of professionals that transcend simultaneous fields of the art of re-mediation (=media interconnection), the establishment of pro-active interpersonal relantionships (= public/audience selection), a conception or use of technological platforms to process and provide data (= databases recurrence), and mainly, an intellectual and informative capacity to curate in a unique and different way (= value-adding). All these sets of competencies can be associated with curator-algorithms, in such a way, in which, not only the appropriation of past history and the informative preferences come into play, but principally, the result of these communicator specific actions.

From this perspective, the curator-algorithm will act from two combined vectors: the formal marking included in the publisher (CMS) and the action of a journalist who experiences the connection with environment and society, creating random and personal markers.

What we presume is that Curation in Journalism can take advantage of technological evolution, but it is not the result of determinism. On the contrary, it is a deliberated action of those who take on the roles of re-mediation in society; they can be Activists, Bloggers, Researchers, Journalists or Communicators.

We have been talking, therefore, about a specialized curatorial action. And that makes sense, mainly, when considering the process of an increasingly noticeable loss of content diversity offered through Traditional Means of Communication in digital spaces. Boczkowski (2010) indicates a concerning reality about this problem: there is a growth in a proportion of "generic" contents shared by the main Means of Communication (and even among third-party providers, such as news agencies); at the same time, there is a reduction of independent media companies in journalism conglomerates to perform it's "guard dog" role in an active way, resulting in "more of the same" informative terms in digital sites which offer themselves to keep their readers well informed about world events.

> The leap of mechanical reproduction to the digital one, whether it is about news or other areas of the symbolic work, introduced, in the beginning of the twenty-first century an information abundance age marked by the concern over the increasing cost-effectiveness between information quality and quantity. Having more information available than resources to process it and the fear that this growth in volume causes the resulting product deterioration and consumption experiences are not new, however, they have been more evident during these last years. (Boczkowski, 2010, loc.2445, our translation)

It is clear that not even every curator (in the etymological sense of the word) is a potential communicator in the digital world, or a journalist. We also observe that the communicational action in the digital net is even more anchored in curatorial processes, which range from simple recommendation systems to complex algorithm construction, passing all of them through inherent mediation to the communicational activity.

In *Facts are Sacred: The power of data,* edited by the British Newspaper The Guardian, Rogers (2011) discloses that a curation performed by an expert (in this case, a journalist) predicts distinct skills, however: also includes hours compiling and analyzing data in Excel tables and PDF documents available on the Internet, looking for an informative standard and with news value to, consequently, create a perspectival informative scene in a new way (2011,loc. 64) .

The same author states that digital data abundance transforms Journalism and also, Communication. This, so called, data journalism becomes, most of the time, curation, as stated below:

> **Has Data Journalism become Curation? Yes, sometimes it has.** There is, today, a certain data quantity available in the world that we aim to offer in each news story the main facts - and finding accurate information can become a journalistic activity as intense as searching for the best interviewees (...) **Anyone can do that...** Especially through free tools like Google Fusion Tables, Many Eyes, Google Charts or Timetric – and you can access postings of readers in your Flickr group (...) **However, the most important task is to think about the data obtained as a journalist than as a analyst.** What is interesting about this data? What is new? What would happen if I mixed it with new data? The answer to these questions is of the uppermost importance. It works if we think of a combination thereof. (Rogers, 2011, loc. 56-71, our translation and emphasis added)

Steven Rosenbaum (2011) perceives a broadening curation in society, mixing citizens, professionals, tools and platforms, without any distinction of function or need. According to him, the web and it's users auto-format themselves. In this line of thought we must disagree with the non-consideration of possibilities of communicational exploration of this curatorial process which, today, is available on the net.

Beiguelman (2011, online) considers the information curator someone who uses tools which are associated with filters and platforms that facilitate the activity of agency, generating a distributed intelligence.

From the scenario described here it is possible to affirm that the journalist, especially, and the Communication Field constitute themselves in the most propicious scene to activities of information curation through their own characteristics. It is also evident that the curation action leveraged by tools based on algorithms which would work as communicator partners. The point that is taken into consideration, at last, is about the status of such a partnership and its future perspectives, widening even more an arena for more debates.

Final Remarks

We aimed to, throughout this paper, organize the main ideas which have been thouroughly discussed about the influence of the third web wave in Journalistic Information and Communication Fields as a whole. A scene which we consider typical of the current Cyberculture.

For this purpose, it was necessary to present and explain how the SW, Algorithms and the apps work – founding technical elements to the existence of this recent technical and informational layer, also known as Web 3.0.

Afterwards, we tried to correlate these elements with the process of producing informative contents and communicative actions as a whole, indicating the potential changes. Finally, we discuss the proposition that the communicator, in general, and the journalist, in particular, are heading for the absorption of another role in their performance, outlined by technical and interactive features. All of them due to the scenario where the machine can take on a more active role in the process of producing society's knowledge and opinion on the net.

Evidently, the discussion presented here is complex and, at this moment, not very conclusive. The Cybercultural Scene, which the informative world has entered, will require, again, a cultural and behavioral repositioning from both communication professionals and informative companies. The introduction of the curation activity/role may persist for some years. The context will yet require significant investments in CMS systems, ontologies and controlled vocabularies, in other words, in indirect areas of the core business whose feedback is not immediate. Consequently, it will require rethinking the business models in which pre-web 3.0 aspects are still discussed these days. In this scenario, the word (re)valuation emerges as an aggregation point and as essential raw material for the informative practice in the Semantic World.

Again, a scenario of uncertainties? It is something to be debated. At least the digital world of Communication seems like being familiar with the uncertainty since the early days of networking, being incorporated into its routine. However, it is certainly a scene of extreme value for the world and, consequently, of its artificers - journalists and communicators - that now face the perspective of consolidating their roles in the mechanical digital society that seems to be emerging.

References

ADOLPHS, Peter, Xiwen Cheng, Tina Klüwer, Hans Uszkoreit, and Feiyu Xu. "Question Answering Biographic Information and Social Network Powered by the Semantic Web." In Proceedings of the Seventh Conference on International Language Resources and Evaluation, edited by Nicoletta (Conference Chair) Calzolari, 2764-2768. Valletta, Malta: European Language Resources Association (ELRA), 2010. http://www.lrec-

conf.org/proceedings/lrec2010/summaries/611.html.

ANDERSON, C.W. Deliberative, Agonistic, and Algorithmic Audiences: Journalism's Vision of its Public in an Age of Audience Transparency. International Journal of Communication 5 (2011), 529–547. ISBN 1932–8036/20110529.

ANGELETOU, Sofia, Marta Sabou, Lucia Specia, and Enrico Motta. "Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report." In Workshop: Bridging the Gap between Semantic Web and Web 2.0, European Semantic Web Conference, 93, 2007. http://kmi.open.ac.uk/people/marta/papers/semnet2007.pdf.

AROYO, Lora, and Geert-jan Houben. "User modeling and adaptive Semantic Web." Semantic Web Journal 1 (2010): 105-110. http://www.semantic-web-journal.net/content/user-modeling-and-adaptive-semantic-web.

AUER, Sören, and Jens Lehman. "Creating knowledge out of interlinked data." Semantic Web Journal 1 (2010): 97-104.

BANNIER, Stijn, and Chris Vleugels. "How Web 3.0 Combines User-Generated and Machine-Generated Content." In Transforming Culture in The Digital Age - International conference in Tartu, edited by Agnes Aljas, Raivo Kelomees, Marin Laak, Pille Pruulmann-Vengerfeldt, Tiina Randviir, Pille Runnel, Maarja Savan, Jaak Tomberg, and Piret Viires, 69-77. Tartu: Estonian National Museum, Estonian Literary Museum, University of Tartu, 2010.

BASULTO, Dominic. Are Newspapers Civic Institutions or Algorithms?

BBC News. When algorithms control the world, BBC News, Availabe at http://www.bbc.co.uk/news/technology-14306146 Accessed 14 February 2012.

BEIGUELMAN, Gisele. Curadoria de informacão. Palestra, ECA-USP, 2011. Availabe at http://www.slideshare.net/gbeiguelman/curadoria-informacao. 2011.

BEIGUELMAN, Giselle. "O fim do virtual." Revista Select. São Paulo, SP, 2011.

---. "Curadoria de Informação". Encontros com o Futuro ECA-USP. 21/06/2011. http://grupo-ecausp.com/

BERNERS-LEE, Tim, James Hendler, and Ora Lassila. "The Semantic Web." Scientific American 284, no. 5 (May 2001): 34-43. http://www.nature.com/doifinder/10.1038/scientificamerican0501-34.

BERNERS-LEE, Tim. "Levels of Abstraction: Net, Web, Graph." W3.org, 2010. http://www.w3.org/DesignIssues/Abstractions.html.

BERTOCCHI, Daniela. Narrativas jornalísticas no contexto da web semântica.. In: Records of the 2nd Seminar of Cyberjournalism of Mato Grosso do Sul, 2010.

BOCZKOWSKI, P. News at work - Imitation in an age of information abundance. The University of Chicago Press, Ltd., London . Kindle Edition, 2010.

BOURENANE, Amel, Samuel Szoniecky, and Imad Saleh. "Les perspectives du tagging sémantique", 2009. http://www.ieml.org/spip.php?article154.

---. "The Future of Journalism: Artificial Intelligence And Digital Identities." In Journalism Research & Education Section, The International Association For Media and Communication Research (IAMCR), 1-31. Braga, Portugal, 2011.

BRESLIN, J., Passant, A. & Decker, S.The Social Semantic Web, Heidelberg: Springer-Verlag, 2009.

CARR, Nicholas. Nicholas Carr: 2012 will bring the appification of media. Nielan Journalism Lab. Available at http://www.niemanlab.org/2011/12/nicholas-carr-2012-will-bring-the-appification-of-media/ Accessed: 4 March 2012.

DI LUCCIO, Flavia. "Do Iluminismo à Web Semântica". Pontifícia Universidade Católica do Rio de Janeiro, 2010.

GABLER, N. The Elusive Big Idea. The New York Times. 2011. Available at: http://www.nytimes.com/2011/08/14/opinion/sunday/the-elusive-big-idea.html Accessed: 14 March 2012.

HITZLER, Pascal, and Frank van Harmelen. "A reasonable Semantic Web." Semantic Web Journal 1 (2010): 39-44. http://knoesis.wright.edu/faculty/pascal/resources/publications/HH-ARSW.pdf.

JOHNSON, C. The Information Diet: A Case for Conscious Consumption. O'Reilly Media, 2012. Kindle Edition.

KING, Brandy E. & Reinold, Kathy. Finding the Concept, Not Just the Word. A librarian's guide to ontologies and semantics. Oxford, UK: Chandos, 2008.

LATAR, Noam Lemelshtrich, and David Nordfors. "Digital Identities and Journalism Content - How Artificial Intelligence and Journalism May Co-Develop and Why Society Should Care." Innovation Journalism 6, no. 7 (2009): 47.

LAURENTIZ, Silvia. "Tags e metatags? De Ted Nelson a Tim Berners-Lee." Revista Porto Arte. Porto Alegre, May 2010.

LÉVY, Pierre. A ideografia dinâmica: rumo a uma imaginação artificial? Editora Loyola, 1998.

---. "From Social Computing to Reflexive Collective intelligence", 2009. http://www.ieml.org/IMG/pdf/2009-Levy-IEML.pdf.

LOPEs, Ilza Leite. "Uso das linguagens controlada e natural em bases de dados: revisão da literatura." Ciência da Informação 31, no. 1 (January 2002): 41-52. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000100005&lng=pt&nrm=iso&tlng=pt.

PARISER, E. The Filter Bubble: What the Internet Is Hiding from You. Publisher: Penguin Press HC, 2011. Kindle Edition.

RABAINO, Lauren. 5 new year's resolutions for newsrooms, 30/12/2011. 10.000 Words. Availabe at http://www.mediabistro.com/10000words/five-new-years-resolutions-for-newsrooms_b9610 Accessed: 4 March 2012.

ROGERS, S. Facts are sacred - The power of data. The Guardian Books, London, Kindle Edition, 2011.

ROSENBAUN, Steven. Curation Nation – How To Win In a World Where Consumers are Creators. McGrawHill, 2011.

SIEGEL, David. Pull – The power of Semantic Web to transform your business. Portfolio Hardcover, 2010.

SKIENA, S.S. The Algorithm Design Manual, 2nd Edition, Springer-Veriag London, 2008.

SONDERMAN, Jeff. Three trends from 2011 that will reshape digital news in 2012. Poynter, 30/12/2011. Available at http://www.poynter.org/latest-news/top-stories/157639/three-trends-from-2011-that-will-reshape-digital-news-in-2012/ Accessed: 4 March 2012.

TEIXEIRA, João de Fernandes. A mente pós-evolutiva – A filosofia da mente no universo do silício. Petrópolis, RJ: Ed. Vozes, 2010.

---. Inteligência Artificial: uma odisséia da mente. São Paulo, SP: Editora P, 2009.

---. "O que é inteligência artificial". São Paulo, SP, 1994. http://www.filosofiadamente.org/images/stories/pdf/o_que_e_inteligencia_artificial.pdf.

TURKLE, S. Alone Together: Why We Expect More from Technology and Less from Each Other. Basic Books, 2011.

WEINBERG, D. Too Big to Know. Basic Book. Kindle Edition, 2012.

WORLD WIDE WERB CONSORTIUM, 2004. 'OWL Web Ontology Language: Use Cases and Requirements". W3C. Available at http://www.w3.org/TR/2004/REC-webont-req-20040210/#onto-def.