Freedom and the Internet: empowering citizens and addressing the transparency gap in search engines

Fernando Galindo[1] and Javier Garcia Marco[2]

ABSTRACT

This work contemplates the limits and possibilities of exercising the right to freedom through the use of the Internet. Freedom can be defined as the preservation of the right of autonomy in the daily life of citizens or members of social and political organisations, whilst respecting the utilisation of this right, by oneself, or by one or more persons or citizens. A number of strengths and weaknesses are identified in this regard. The paper examines the way in which search engines like Google exemplify restrictions on freedom: they are enhanced by the use of technical resources that are aimed at the most efficient exploitation of the information available on the Internet; the resources are not utilised to reinforce the rights of the users. Finally, it is argued that the limits imposed on freedom can be overcome with the aid of technical tools such as thesauri that can produce a positive relationship between freedom and Internet.

Keywords: freedom; autonomy; rights; the Internet; open data; search engines; thesaurus

1. INTRODUCTION

Mark Levene (2010, p. XIV) takes the following approach to Web search engines: "Searching and navigating the web have become part of our daily online lives. Web browsers and the standard navigation tools embedded in them provide a showcase of successful software technology with a global user-base that has changed the way in which we search for and interact with information. Search engine technology has become ubiquitous, providing a standard interface to the endless amount of information that the web contains."

The constitution, the risk of automatism and the generalised use of Web search engines brings benefits in access to the content of the Internet but may also lead to serious problems in terms of privacy and transparency and lack of control on the part of the user. More specifically, the keywords of each query and the related metadata may provide anyone who has access to the logs with sensitive information about the users - their behaviours, habits, interests, religious views, sexual orientation, etc. Even worse, some query contents can contain identifiers and quasi-identifiers which may allow the linking of a particular query with a real person[3]. In addition, users have no way of knowing how their query results are obtained. Tips are only given when the search engine administrators want to clarify a problematic search statement. Users can make their own evaluation as to whether the search results are relevant to their needs, but they cannot check what the automatism filters, transform or leave. This means that there is a clear transparency gap and an impairment of the autonomy of the user, and this leads us to a discussion on the link between the Internet and freedom, understood as the preservation of the right of autonomy in the daily life of citizens or members of social and political organisations. The debate focuses on the connection, lack of connection or disconnection between two basic phenomena that are protected by the rule of law: i) the exercise of the right to freedom; and ii) the practices and uses of the Internet. In recent years the importance of the Internet has been augmented by the generalised use of the mobile telephone as a habitual method of communication and as social sensor[4].

The aim of this work is to consider how fundamental human rights are respected by the information that is stored and by the characteristics of an individual's access to the Internet. The following citation can be considered as a reference; it establishes a basis for all rights that are integrated into the legal systems of European Union countries: "Conscious of its spiritual and moral heritage, the Union is founded on the indivisible, universal values of human dignity, freedom, equality and solidarity; it is based on the principles of democracy and the rule of law. It places the individual at the heart of its activities, by establishing the citizenship of the Union and by creating an area of freedom, security and justice".[5]

This wide ranging expression of freedom is not only found in legal texts. The texts simply demonstrate the fact that the principle is generally recognised as central to the organisation and functioning of democratic societies and their judicial systems. As the philosopher Axel Honneth (2014, p. 93) has commented with regards to the principles and values of our society "...in modern societies, there is but one single value that forms the basis for the legitimisation of the social order: for the different types of systems of action of this class of society it may be enough that they specifically embody, in their functions, the ethical idea that contributes to all subjects achieving an equal measure of individual freedom".

In a manner that is coherent with the universal acceptance that the principle of freedom belongs to, or is characteristic of, democratic societies, we intend to examine if it is possible to establish a solid relationship between freedom and the Internet[6]. At the same time, we will

answer the question: 'Can the Internet enhance the exercising of the right to freedom, avoiding any type of coercion to the principle of the autonomy of the will, as the basic element of freedom in a democracy?'

The definitive response to this question will be provided at the end of this work, which is structured as follows: Section 2 looks at some clear cases of control of the use of the Internet and the freedom that (according to others) it facilitates; Section 3 deals with the Law, the effective regulation of the information that, in Europe at least, is organised through the Internet in observance of the regulations on the protection of personal data and open government[7]; Section 4 discusses the fact that limitations on Internet use also exist with regards to the functioning of the access techniques themselves: the use of search engines is not transparent and, in order to satisfy the demands of regulation and freedom, this requires techniques that are most suitable for organising and accessing information, such as thesauri; finally, and by way of a conclusion, Section 5 will respond to the question of whether freedom, the autonomy of the will can be fostered by the use of an adequate organisation of information and its storage and access in the Internet.

2. THE CONTROL OF FREEDOM, UNDERSTOOD AS THE AUTONOMY OF THE WILL

This section suggests that in spite of the arguments that the Internet is an instrument that fosters and encourages freedom, there are, in practice, cases in which we can see the control of the Internet and freedom as the autonomy of the will that others claim it facilitates.

The first point of reference is freedom: a wide ranging expression that has been refined, over the years, by a variety of thinkers and philosophers. During the Renaissance, René Descartes (1637, pp. 11-16) saw freedom as the liberty of 'understanding', dependent on the individual and not on an external authority. For Kant (1797, pp. 356-372), in the 18th century, the freedom of the citizen was the freedom to work: the ability "to obey no law other than the law to which the citizen has given their consent". Also in the 18th century, Rousseau argued that (1762, p.4) "*L'homme est né libre, et partout il est dans les fers*"– man is controlled and in order to be free, he must only obey himself, through the implementation of a mechanism of the general will, dedicated to the creation of the law that represents that general will.

As we will see, the relationship between the Internet and the above mentioned notions of freedom is concerned with the fact that the Internet is not simply a network of computers and communication that expands knowledge; rather, through the storing of information, it can be a channel or means for limiting the expression of 'understanding' or thought, as occurs in a number of countries (2.1).

The Internet affects freedom to work when, in line with the content of the 'Internet of Things', it becomes a cybernetic instrument of control through the transmission and execution of orders emitted by people or instructions generated by processes and directives programmed to such effect, following the corresponding routine or algorithm (2.2).

Finally, the Internet is also an instrument of political control: it allows for remote monitoring through cameras and sensors, of the activities and processes that occur in any and all social environments (2.3).

After consideration of the above mentioned issues, we will be able to draw some initial conclusions with regards to freedom and the Internet (2.4).

2.1 THE LIMITATION OF THE FREEDOM OF EXPRESSION AND THE INTERNET.

The phenomena of the limitation of freedom of expression that is produced by the Internet is so evident that it is even accepted by those who believe that the Internet foments the possibility of the dissemination of sentiments and opinion by many more people than traditional communication media. They recognise that generalisations have to be refined and that the issue, like many other legal questions, is complex and multifaceted. In a collective work edited by UNESCO (2011, p. 74), Dutton argues that the relationship between freedom of expression and the Internet is relative, as both concepts must consider other rights:

"Protecting certain human rights or freedoms often has a direct and immediate impact on other rights and freedoms. Thus, the preservation of one freedom can limit another. Balancing these conflicting values and interests is only likely to be a resolved through negotiation and legal-regulatory analyses. This will probably vary cross-nationally, if not locally. Resolution of these balancing issues requires a broad view of the larger ecology of policies and regulations shaping freedom of expression."

2.2 ECONOMIC LIMITS ON THE USE OF THE INTERNET

One basic limit on the exercise of freedom of action through the Internet is economic; if an individual does not have the resources to access the Internet then they are, obviously, unable to undertake any kind of action. Access to the Internet implies access to an enabled device and the ability to pay for the connection in order to take advantage of emails and other means of disseminating information. Therefore, the Internet is not the same as freedom of action and communication for everyone, this right is limited to those that possess sufficient resources.[8]

A further basic limit is having the necessary technical knowledge and skills to utilise the Internet. However, this limitation is less and less relevant as education and training means that more and more people are familiar with the requirements of participation in the Knowledge Society. State organisations and companies are supplying their citizens and workers with training in practical computer skills and higher level users and the popularity of mobile phones throughout the world have generalised the assimilation of technical knowledge.

2.3 POLITICAL FREEDOM

Political freedom is limited by the relative ease with which controls can be imposed on the use of the Internet by both public and private bodies. Two obvious examples of countries that use these controls to limit freedom of political expression are China and Turkey who tightly regulate Internet content. There are regulations that allow government organs to censor the content of what can be put on the net and what can be seen.

In the private sector, in all countries, suppliers of services can emit, cut or censor content at the same time as they obtain information about the user (address, personal details, purchase history, messages etc.). On a more serious note, it is important to be aware of the way that companies that offer Internet services and content surreptitiously manage the behaviour of the users, thereby reducing their effective freedom. This occurs, for example, when only a part of the relevant information is supplied or methods that attract or direct attention are employed. This issue is most relevant to search engines; they should develop and implement a code of ethics in the same way as other institutions that provide information and references, such as libraries and official archives.

All this means that the Internet can be used to both limit and strengthen freedoms, it is therefore impossible to argue that freedom and the Internet are equivalent, or unequivocally linked[9].

Furthermore, it has been shown (for example, in the United States) that without the need for official censorship to limit the publication of information, the intelligence agencies control and monitor Internet communication, collecting and analysing information, and, in some cases, this even leads to punishment. Surveillance can be undertaken without any evidence of wrongdoing or possible danger in communication between individuals. In other words, this is a control that has not been approved by a judicial authority that in a democratic country is charged with intervening in situations in which there is sufficient reason to investigate and sanction in order to prevent a potential criminal act.[10]

2.4 THE INTERNET ALLOWS FOR THE OBSERVATION AND CONTROL OF INFORMATION AND COMMUNICATION

The previously mentioned examples illustrate the well-known fact that the Internet allows for the observation and control of the information and communication that is generated between the users of computers/mobile phones. It is a control that can be utilised as and when required and it means that the Internet can limit freedom as it restricts the exercise of free will, even when the principles of free will are respected by the users (control is exercised by those that control the Web, not those that use it).[11]

It is unarguable that through social networks such as Facebook, Twitter, Instagram, Linkedin, Youtube, blogs etc. people can publicise opinions, send videos and 'freely' supply information; however, this does not alter the fact that those that are responsible for the Internet can modify the transmission of that information, they have the resources to intervene in the publication of information as they see fit.

This leads us to a paradox: as will be seen in the following section, the Internet is subject to a series of complex regulations concerning the information that is stored and available; regulations that are aimed at fostering freedom and avoiding abuse.

3. THE REGULATION OF THE INFORMATION CONTENT OF THE WEB ACCESSED THROUGH SEARCH ENGINES

This section deals with the law and the regulation of the information content of the Web. It is compulsory for the Web search engines to respect the regulation of the information and communication technologies that is summarised in this paragraph: the regulation establishes that the Internet should be subjugated to the right to the exercise of freedom. More specifically, the law on freedom is exemplified by the satisfaction of the requisite of express consent for the use of personal data or the exercise of the principle of informational self-determination, elements of the law on protection of personal data and European constitutions that must be respected by the structure of the information that is accessible by means of the Internet (3.1).

This is coherent with the 'open data' movement which aims to make all information available to Internet users. Supporters would argue that their proposals came about as a consequence of equating the Internet with freedom, but this can only occur if the information is made available to everyone in compliance with the legal regulations to which access is subject. Here, we are referring to the law that must be respected in its various forms (copyright, patents, intellectual property, etc. in addition to the law on protection of personal data) in order to prevent abuse (3.2).

In relation to the right of freedom, the same thing happens with the idea of 'open government'. As we will see in subsection 3.3, the initiatives of open data and open government can only be implemented when they respect current legislation and the principles of, for example, the separation of powers, transparency, protection of personal data and democratic participation, etc.

3.1 THE JUDICIAL FRAMEWORK CONCERNING STORAGE, RECOVERY AND SEARCH FOR INFORMATION

Before the existence of the Internet computers were used for storing and processing personal information in an arbitrary manner, or through the use of closed communication networks. With the expanded use of Information and Communication Technologies in business, commerce and government, it was noticed that citizens' rights to freedom were being violated by the owners of computer systems, especially those responsible for managing databases which held personal information for specific ends, such as purchasing products, the supply of medical services, insurance policies etc. It was soon recognised that the exercise of personal freedom by the holders of the personal data (who gave consent for the use of their information for specific ends) was vulnerable. The personal data could be used by the receivers and those responsible for the databases for other means, without having obtained further consent, that is to say, the exercising of free will on the part of the individuals concerned.

The first legal cases on this issue took place in the United States. The judgements took into consideration precedents for intervention based on Anglo-Saxon law and sentences dictated by the United States at the beginning of the twentieth century on the 'privacy' of accused criminals whose photographic images were published without their consent.[12]

In Europe, this approach was seen as rather strange, for the following reasons:

i.   It makes reference to 'privacy', a term that is not commonly used in continental judicial systems. A more suitable expression, from this perspective, would have been 'intimacy'[13].

ii. It would be better to make use of 'the expression of freedom' – the exercise of freedom through the recourse of consent.

iii. The judge handed down a sentence that was not based on any previously enacted law on the issue (this was perfectly legal in the United States).

However, these complications were not too problematic. The sentence was coherent with the Common Law of the United States. For a judge to intervene there has to be a precedent, a previous sentence on a similar case. In this instance, the precedent was the definition of privacy as an individual right that had been violated.

Given the same potential infringements of personal freedoms, following the standards established by the democratic constitutions, European courts adopted a different approach to that of the United States[14]. General regulations were passed for the protection of personal data, aimed at the preventing the above mentioned infringements and avoiding attacks on the right to freedom/privacy. Each country was to establish an independent administrative authority – the Data Protection Office – to which authorities and bodies responsible for the programming, implementation and management of databases were required to declare the general characteristics of their personal data files and the way that the data would be utilised. At the same time, procedures were introduced that enabled interested parties to consult the databases with regards to their own personal details and to check if the information was being used in accordance with their express consent. If companies or authorities did not allow this, citizens could use their right to 'informational self-determination' (the right to data protection/freedom/privacy) and call on the Data Protection Office to use their legal powers to force the companies or authorities to respond to interested parties, if they did not, the Office could impose a sanction.

Since then, the regulation of freedom, through the law on protection of personal data has evolved throughout the European Union. In the United States, on the other hand, in line with their legal tradition, Europe's preventative mechanisms do not exist. The judges are responsible for deciding if there has been a violation of 'privacy', based on precedent and as a consequence, the development of the Internet, its programs and applications have been very different on the other side of the Atlantic. In Europe, there are specific organs and procedures for the protection of personal data which give citizens the right to defend themselves against the violation of their freedoms; this is not the case in the USA[15]. This fact has had an impact on the development of the Internet in USA, the programs, applications and services have not been created in accordance with the practices and regulations on the protection of personal data that are in force in Europe.

This fact prejudices companies that adapt programs to European needs without knowing how to ensure that they avoid sanctions by users or other companies for not complying with legislation on protection of personal data. This, in itself, supposes a limitation of freedom as failure to comply with the regulations or the principle of informational self-determination and the use of personal data without express consent is an infringement of current European regulations.

From this perspective, the cultural tradition of the USA with regards to the Internet can cause problems in other countries that have a different approach to the question of limitations on

freedom and rights. It was in Europe that legislation was first introduced for the protection of freedom and use of personal data by companies and organisations working with information and communication technologies. Some of the practical consequences of these regulations have already been mentioned, but there are others, for example, the 'right to be forgotten' or the deletion of information, that have been recognised by European courts in relation to Google[16]. This type of contravention of regulations on data protection was foreseen by the European Parliament and the European Council and will feature the forthcoming changes to regulations[17]. A further example is the exercise of the right to data protection, which is subject to regulation in Spain, as in other European countries[18], but it is difficult for companies to put the rules into practice when applications are developed without taking into account the existence of institutions dedicated to the protection of personal data, as in Europe.

The risks are self-evident: those who access information on the Internet are making use of a daily resource that is close to illegality if not used correctly, that is to say, anonymously. Here, we refer to the use of mechanisms such as 'cookies', programs that gather personal information about users of websites. These tools are used by search engines, but also by radio stations and TV or even organs of the state. They are generally employed to explore possibilities for new business techniques with advertising and basic product placement at the same time as sending messages to the users via social networks (Facebook, Twitter etc.). By means of the 'cookies', all kinds of information is gathered and exploited for commercial ends. The communication media make the best defence that they can with regard to the use of information that is published on the Internet; this is the case with search engines that offer information without compensating the sources for the use of that information. In Europe, there are rules that regulate these abuses of freedom.

The following subsections examine European Internet regulation on problems related to 'open data' and 'open government'.[19] There is a risk that the implementation of the idea of freedom of expression (the opinions, proposals and details that are 'open' to Internet users) may be problematic due to the way that this information may be exploited by third parties. This is the central theme of subsections 3.2, 'Open data', and 3.3, 'Open government'.

3.2 OPEN DATA[20]

This section focuses on warning against the technologists' arguments that the Internet is the same as freedom. They suggest that the information and data available on the net are 'open data', of free use and outside the rights of ownership (intellectual/industrial/personal property etc.).[21]

This idea contradicts the concept of freedom that is written in the European Constitution and the Constitutions of all democratic countries. As, for example, those responsible for the communication media well know, data and information is never free; it is always linked to someone. In other words, there are always rights and obligations, whoever publishes whatever is published, and this is also true for the programs that manage the information. Paradoxically, this is an argument that is used and defended by the same technologists who believe that all data and information is free.

According to Article 6 of the European Constitution, freedom is: "The right to liberty and security. Everyone has the right to liberty and security of person; (Article 7): Respect for private and family life. Everyone has the right to respect for his or her private and family life, home and communications; (Article 8): Protection of personal data. Everyone

has the right to the protection of personal data concerning themselves, such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning themselves, and the right to have it rectified. Compliance with these rules shall be subject to control by an independent authority"[22]. In other words, the exercise of freedom must respect the freedom of others and 'the rights of others'.

The concept of open data restricts the correlation between freedom and the Internet. In this case, it is the erroneous understanding of freedom of opinion that is offered by the Internet. It is a limit on freedom itself because those in favour violate the rights of those that publish information on the Internet, appropriating that information for uses and ends whose objectives and reach have not been expressly consented to.

The problem grows when open data is accessed through search engines like Google, that are not transparent and made up of mathematical formulas that are unknown to users who themselves have no right to access them. This is an issue that will be dealt with in section 4 of this work. In the case of search engines, European law may be infringed by:

1. The technicians that believe that the data is open (when it is not) and make it accessible through the net.
2. The institutions responsible for safeguarding the legality of whether the data are open or not.
3. The companies that own the search engines and the algorithm that accesses and supplies information.
4. Individuals that access information and use it without considering its legality.

3.3 OPEN GOVERNMENT

Technicians who support the 'open data' movement also demand that the public administrations make available the data and information related to their functioning and decision making, this is known as 'open government'.

They argue that open data and open government will increase the effectiveness of the Internet as governments will have to publish data on their service provision: unemployment benefit; health services; the granting building licences; statistical information; public registration services; budget management; public expenditure; civil service salaries etc.

This is the data which the public administrations use when making decisions. The technicians demand the publication of this data so that it is freely available to citizens and companies. With this data, new programs and applications could be developed with aims that are very different from those for which the data was collected and managed.

To some extent, this process is already taking place, in line with the democratic obligations of governments and administrations with regards to the efficiency of the management of public funds and the demands of the citizens for greater transparency.

Furthermore, at least in Europe, governments take precautions when they publish data, in an attempt to safeguard the rights of citizens who have given consent for their data to be used

for specific ends. If this were not the case, the governments, administrations and civil servants could be guilty of violating the freedom and the rights of the individual.

It is logical, therefore, that the administrations demand that the use of the open data by programmers/technicians/companies does not imply any cost, or, if there are any costs for the administrations that supply and program the data with guarantees that safeguard the personal details of the citizens, those costs are met by the organisations or individuals that wish to make use of the information.

This activity becomes problematic if the only tools for accessing information are the search engines that offer no context for the data that is supplied. The following section contemplates the advantages and disadvantages of this situation.

4. SEARCH ENGINES AND THESAURI

In this section, we will see that the limitations imposed on freedom by the use of the Internet are not mitigated by simply complying with current regulations on the information that is made available to users. There are limitations that are concerned with the functioning of the methods of access; programs and applications of generalised use, such as Google infringe the law due to their lack of transparency and the way that they work. Subsection 4.1 focuses on the consequences of the aforementioned techniques while subsection 4.2 reveals that, in spite of everything, there are other methods of access to information, (for example, thesauri) that can complement Google and other similar search engines in ways that are more compatible with citizens' rights. In contrast to the opaque functioning of search engines like Google, thesaurus standards (ISO 25964-1:2011, ISO 25964-2:2013) can provide a transparent way to disclose the search procedure[23]. This approach might contribute to bridging the gap between the Internet and current legislation, strengthening the power of citizens' access to information and overcoming obstacles to freedom inherent in the use of the Internet[24].

4.1 INTERNET SEARCHING

The world of references and Internet searching is increasingly dominated by Google and other new generation search engines. Google is the market leader and has a global share of 66.74%, Bing, Yahoo and Baidu claim around 10% each (Net Applications, 2015). Even in academic contexts, simple keyword Google searches are increasingly preferred by students (Georgas, 2013, 2014, 2015) and teachers (Kemman, Kleppe and Scagliola, 2014). Not only do users prefer Google when searching the Internet, they also prefer to use it in the simplest possible way. The vast majority of searches do not go beyond the keywords that naturally come to mind; users do not examine the concept carefully to find synonyms or related terms, they do not use commands, do not expand or refine their searches, do not examine the metadata and they do not look at results other than those which appear on the first page.

Among the problems that partially obscure the excellence of many Internet search engines (both general and specialised) is their lack of transparency and the powerless dependency that they generate among users. When analysing improper results that can be harmful to users, it has been noted that: "Search engines lack any transparency to clarify how results were found and how they are connected to the search terms." (Machill, Neuberger and Schindler, 2003). Kemman, Kleppe and Scagliola (2014) pointed out that even scholars are becoming

increasingly dependent on "black boxed algorithms", calling into question the academic principles of provenance and context. Search comfort and efficiency are certainly positive values, but they are not the only ones to be considered from a long-term perspective.

Transparency problems are even more evident in business situations where consumers are involved, and consumer protection agencies and governmental bodies are increasingly aware of this (ECME and Deloitte, 2015). Lawyers and advocates of protection of privacy are now going further than aspects that purely relate to privacy and are focusing their attention on "algorithm transparency"; an example would be the case of Marc Rotenberg, president of the Electronic Privacy Information Centre (Unesco, 2015).

Therefore, an initial conclusion would be that the Internet is not the same as freedom; its basic methods of functioning, in particular, the global use of search engines, does not respect the law on protection of personal data which is central to the right to freedom and informative self-determination, as written in European legislation.

4.2 FINDING A PRACTICAL STRATEGY FOR WORKING WITH SEARCH ENGINES

We now come to the problem of "algorithm transparency" and its private and public consequences. Industrial secrecy is, of course, a key question in this area. Open policies on the publication of search algorithms are not to be expected until the field becomes commoditized; there is more room for openness with networks of concepts that support semantic searches, mainly because open sources are already being used for this purpose. But far from general strategies that are currently difficult to plan and carry out, some small steps could be taken that would be very beneficial for the search engine firms themselves, deploying a gradual approach without threatening their competitive advantages.

The best candidate for a first step would be equivalent terms. In the current situation, users can rarely be sure that the synonym strings that empower their searches include (or not) the terms that they would find relevant. It would be as simple as including an option in the search menu for consulting the list or a dropdown menu from a selected term. Users could even propose new synonyms or discuss current ones, contributing to the improvement of the tool, as for example, with Google Translator.

Later, this strategy could be extended to knowledge graphs, disclosing the open access sources and the relations among their concepts. In the case of associative and hierarchical relations, the problem of successfully communicating with the users increases.

Regarding the communication of the model, good tutorials exist, so the problem is much less concerned with information literacy than advertising the model to wider audiences. If the search engines could be co-opted, because they have shown how the thesaurus model can be a beneficial tool for increasing user transparency and feeding knowledge graphs with users' suggestions, they can do much of this work themselves.

As search engines are becoming increasingly 'semantic', there is also a need to improve the communication of these semantics to their users. The thesaurus model seems a good candidate for this purpose (García Marco, 2016).

So, search engine firms and consumer protection officials and advocates working in the increasingly important economic field of web retrieval and advertising could find, in thesauri, a model to increase transparency and to, at least, empower, the growing minority that is becoming more vocal in defending their rights; people who are increasingly upset by the current opacity of Internet searching and advertising, and are, perhaps, in the process of embracing the largest Internet players.

However, transparency is not only a question of the right to privacy, it is a prerequisite for empowering users and allowing them to make better choices about their search strategies, even giving them the chance to improve the existing search tools. Previous studies show that, though the thesaurus model is clear and understandable for indexers, users find it difficult and unattractive. Nevertheless, when provided with a basic introduction, they prefer to gain more control of their search process by using them (Greenberg, 2004). It would seem that the users-thesauri problem resides in communicating the model and making it more user-friendly in different contexts.

Although thesauri can be no more than instrumental in many of these questions, the model they offer can be very useful for improving the transparency of the search process which involves critical steps for an informed decision, such as providing context and gathering or filtering sources, search terms and results. One of the most relevant aspects of knowledge maps versus search algorithms is transparency. Knowledge maps are open for everybody to discuss; but algorithms are increasingly proprietary and secret, and only the tips about their functioning are provided to the public. Navigating (instead of simply surfing) is all about having maps, and the thesaurus model is a parsimonious and formal way to codify concept maps for information retrieval. These maps can be subsequently offered, in a more understandable way, to users.

Thesauri can be especially useful for solving the transparency gap by storing the concepts, terms and relations that are of the interest of users. But they must be presented in clear and intuitive ways and be easy to understand. For this purpose, the balance that thesauri offer between representational power and simplicity could be decisive. The 'core' thesaurus model is powerful enough to support search expansion and restriction, but, at the same time, it is relatively easy to communicate in a web environment, at least in its main functionalities, by using different metaphors like dropdown-able breadcrumbs, nested folders, graphic maps, Venn diagrams and other visual tools. In fact, some redundancy of codes is needed to accommodate different user cognitive styles and communicative preferences.

With a more user-friendly thesaurus model, new ways to express conceptual relations that could be better understood by users in a web environment must be formally adopted. The arithmetical metaphor that is currently being used to denote the thesauri relations when both terms are presented, 'smaller than' (<) and 'greater than' (>) signs, which are typical in thesaurus presentations, could be complemented by a spatial term that provides buttons for hierarchical navigation when only one of the terms of the relationship is present (as is usual in web interfaces). The upwards and downwards arrows could be used to signify more general and more specific concepts; similarly, the continuous horizontal double arrow could express the relationships between sibling concepts, and the dash could refer to related terms.

Once the mechanics of Internet searches, advertising and their subjacent ontologies become more transparent to the consumers of information, users will want to comment and discuss them. Thesauri and knowledge object solutions in the Internet should include devices to

codify and store these discussions about knowledge maps in the social media and this could be a prospective challenge to be addressed in new editions of the current standards, perhaps by simply including user notes in the model, initially as a special custom note.

5. CONCLUSIONS: WEB SEARCH ENGINES AND THE AUTONOMY OF THE WILL

This work has revealed the manner in which relationships between the Internet and the autonomy of the will involve problems that are difficult to overcome. It has also been noted that these problems are of less import if the use of the Internet, at least in terms of the storage and retrieval of information, takes into account current European regulations.

We have examined the difficulties that result from the most common method of accessing the Internet – the utilisation of search engines that employ algorithms   that are not transparent. We have suggested a solution in a specific case of the search for information: the proposal of a model of information design that respects its content, satisfying this objective by integrating the use of thesauri as a method for presenting information in a manner that is compatible (from the point of view of the user) with the complex requirements of the law.

The relations provided by the thesaurus model are enough to show, in a simple scheme, the synonyms and the polysemy distinctions that have been considered. In this way, users can decide if the search is slanted or if any, relevant terms or relationships are being disregarded. If further automatic query expansion or filtering has been used, hierarchical and related relations could also provide context to users, even in the form of graphic presentations.

Ideally, search engines could provide users with tools for incorporating these terms and relations in a search, so that it is improved. In addition to this, a contact address could be also provided so that users could report problematic assumptions. Not only would users be greatly empowered, but the search engines would get a free mechanism to improve their knowledge graphs and provide users with new functionalities that would contribute to user satisfaction.

These new functionalities should not impair the simplicity of search engine interfaces. Perhaps they should be provided only as optional or complementary features for conscious and professional users that would need them and know how to use them, and also as an occasional tool for normal users that become unsatisfied or suspicious about a specific search.

Using the thesaurus model to enhance search engine transparency could be a solution to a problem of great magnitude: for many users, the search engine *is* the Internet and their access to the Web, which has become the effective information provider. Certainly, users have a degree of freedom, because in principle they can choose between several search engines and Internet portals. But this is not a real freedom: there are more and more data that are of limited use, due to access and legal regulation or the requirement of informed consent, or because the objectives of storage are different than those of use. Besides this, the search engine industry is heavily concentrated, as it has been shown above, and more search transparency is provided only in experimental search engines with a very limited coverage on the Internet.

In this context, the implementation of a solution built on the thesaurus model could be increasingly feasible as search engines rely more and more on semantic graphs, such as those provided by Wikidata. As a result, search engines would become more transparent and respondent, and this would contribute to more autonomous and empowered users in a world where public and personal opinion is increasingly manipulated by BATS and other opaque automatic agents that pose a great danger to our liberties and democracies.

REFERENCES

Aitchison, J. and Dextre Clarke, S. (2004), 'The thesaurus: a historical viewpoint, with a look to the future.' In Roe, S. K. and A. R. Thomas. 2004. The Thesaurus: Review, Renaissance, and Revision. New York: Haworth Press, pp. 5-21.

Descartes, R. (1637), Discours de la Méthode http://classiques.uqac.ca/classiques/Descartes/discours_methode/Discours_methode.pdf

Dextre Clarke, S. G. and M. L. Zeng. (2012), 'From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling (2012)' in *Information standards quarterly* vol. 24(1), pp.20-26.

Dutton, D.H., Dopatka, A., Law, G. &Nash, V. (2011), Freedom of connection, freedom of expression: the changing legal and regulatory ecology shaping the Internet (Paris: UNESCO) http://unesdoc.unesco.org/images/0019/001915/191594e.pdf

ECME Consortium & Deloitte. (2015). Study on the coverage, functioning and consumer use of comparison tools and third-party verification schemes for such tools: Final report prepared by ECME Consortium (in partnership with DELOITTE), EAHC/FWC/2013 85 07. [S. l.]. http://www.konsumenteuropa.se/globalassets/final_report_study_on_comparison_tools.pdf

Galindo, F. (2014), 'La regulación de los datos abiertos' in IBERSID: revista de sistemas de información y documentación, v. 8, pp. 15-18, 2014.

Garcia-Marco, Francisco-Javier. (2016), 'Enhancing the visibility and relevance of thesauri in the web: searching for a 'hub' in the linked data environment' in Knowledge Organization vol 43(3), pp. 193-202.

Georgas, H. (2013). 'Google vs. the Library: Student Preferences and Perceptions When Doing Research Using Google and a Federated Search Tool' in From: portal: Libraries and the Academy vol. 13:2, pp. 165-185. http://muse.jhu.edu/journals/portal_libraries_and_the_academy/summary/v013/13.2.georgas.html

Georgas, H. (2014). 'Google vs. the Library (Part II): Student Search Patterns and Behaviors when Using Google and a Federated Search Tool' in Portal: Libraries and the Academy, vol. 14(4) pp. 503-

532. http://muse.jhu.edu/journals/portal_libraries_and_the_academy/summary/v014/14.4.georgas.html

Georgas, H. (2015) 'Google vs. the Library (Part III): Assessing the Quality of Sources Found by Undergraduates.' in *Portal: Libraries and the Academy* vol 15(1) pp. 133-161. Project MUSE. Web. 26 Sep. 2015. https://muse.jhu.edu/

Greenberg, J. (2004). 'User Comprehension and Searching with Information Retrieval Thesauri' in Cataloging & Classification Quarterly, vol.37(3-4) pp. 103-120. doi:10.1300/J104v37n03_08

Gregory, J. (2013) 'Government Control of the Internet.' Slaw Canada's online legal magazine, 16 January 2013. http://www.slaw.ca/2013/01/16/government-control-of-the-internet/

Honneth, A. (2014), El Derecho de la libertad. Esbozo de una eticidad democrática (Madrid: Katz Editores).

ISO 25964-1:2011. Information and documentation. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Geneva: International Society for Standardization.

ISO 25964-2:2013. Information and documentation. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies Geneva: International Society for Standardization.

Kant, E. (1797), 'Erläuternder Bemerkungen zu den metaphisischen Anfangsgründen der Rechtslehre' in Kants Werke.Akademie Textausgabe,pp. 356-372 (Berlin: Walter de Gruyter & Co 1968)

Kemman, Max, M. Kleppe & S. Scagliola (2014). 'Just Google It.' In C. Mills, M. Pidd & E. Ward. Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities. (Sheffield: HRI Online Publications, 2014.)http://www.hrionline.ac.uk/openbook/chapter/dhc2012-kemman

Koukiadis, D (2014), Reconstituting Internet Normativity (Baden-Baden: Nomos).

Lee, O. (2012), Waiving Our Rights. The personal data collection complex and its threat to privacy and civil liberties (Plymount, UK: Lexington Book).

Levene, M (2010), An Introduction to Search Engines and Web Navigation (Hoboken, Canada: John Wiley and Sons).

Machill, M., Neuberger, C., Schindler, F. (2003). 'Transparency on the Net: functions and deficiencies of Internet search engines' in Info vol. 5(1) pp. 52-74.

Net Applications.com. (2015). 'Desktop Search Engine Market Share', in Netmarketshare.com. Irvine, CA. Net Applications, August 2015. https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0

Pàmies-Estrems, D., Castellà-Roca, J., Viejo. A. (2016). 'Working at the web search engine side to generate privacy-preserving user profiles'. Expert Systems with Applications, Volume 64, 1 December 2016, pp. 523-535.

Rousseau, J.J (1762), Du contrat social, ou principes du droit politique (Amsterdam: Marc Michell).

Saeed, A. (2009), 'Fast Internet access becomes a legal right in Finland',in CNN, international edition, 15 Oct. 2009.
Digitalbiz. http://edition.cnn.com/2009/TECH/10/15/finland.internet.rights/index.html ?iref=24hours

Statista. (2015). 'Worldwide market share of leading search engines from January 2010 to July 2015'in Statista: the Statistics Portal. New York: Statista Inc. http://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

Sussman, L. R. (2000) 'Censor dot gov: the Internet and press freedom 2000'. Journal of Government Information, v. 27, Issue 5, p. 537–545, Sept./Oct. 2000.

Thompson, K.M., Jaeger, P.T., Green Taylor, N., Subramaniam, M. & Bertot, J.C. (2014), Digital Literacy and Digital Inclusion: Information Policy and the Public Library (London: Rowman & Littlefield Publishers).

Unesco. (2015). 'Privacy expert argues "algorithmic transparency" is crucial for online freedoms at UNESCO knowledge café' in Communication and Information Sector. News and In Focus articles, 04-12-2015.

http://www.unesco.org/new/en/communication-andinformation/resources/news-and-infocusarticles/allnews/news/privacy_expert_argues_algorithmic_transparency_is_crucial _for_online_freedoms_at_unesco_knowledge_cafe/#.VoEkMjZnlZ-

Warren, S. D., Brandeis, L.D. (1890). 'The Right to Privacy' in Harvard Law Review, v. 4, n. 5, p. 193-220, dec. 1890.

SOURCES OF FUNDING

FOOTNOTES

[1] Prof. Fernando Galindo, Department of Philosophy of Law, University of Zaragoza, Spain: cfa@unizar.es

[2] Prof. Javier Garcia Marco, Department of Documentation Sciences, University of Zaragoza, Spain: jgarcia@unizar.es

[3] See Pàmies-Estrems, D., Castellà-Roca, J., Viejo. A. (2016), p. 524.

[4] In 2007, 55% of homes in Europe were connected to the Web; by 2014, this figure had increased to 81%. (EUROSTAT: http://www.ontsi.red.es/ontsi/es/indicador/penetracion-telefonia-movil-en-hogares). With regards to individuals that use the Internet on a weekly basis, "...in 2014, the number of users was highest in Holland, Denmark, Luxemburg and Sweden (over 90%), considerably more than the European average of 75%."(EUROSTAT: http://www.ontsi.red.es/ontsi/es/indicador/individuos-que-usan-regularmente-internet).

[5] Preamble to the "The Charter of Fundamental Rights of the Union". http://www.europarl.europa.eu/charter/pdf/text_en.pdf

[6] At first glance, this is what happens: the use of the Internet significantly increases 'freedom of expression' or (essentially the same thing) the field of knowledge, actions and expressions of the individual that uses the Internet in comparison with a person who does not. This is the basis for the debate on 'digital inclusion' that can be seen in Thompson, K.M., Jaeger, P.T., Green Taylor, N., Subramaniam, M. and Bertot, J.C (2014).

[7]A detailed study of the complex regulation of the Internet in Europe (and a reference for this work) can be found in Koukiadis (2014).

[8]In Finland, the right to access to the Internet was established as a right of all citizens in October 2009. In reality, this means that Finnish telecommunication companies are obliged to provide broadband access throughout the country (Saeed, 2009). Obviously, the Finnish people still have to pay for this right.

[9] On the expansion of these limitations, read: Sussman (2000).

[10] This should not be surprising: the abuse of the use of personal data in the USA goes far beyond the military context; it even occurs in the areas of employment selection procedures and applications for credit or loans. For more detailed information, see: Orlan Lee (2012).

[11] For a brief history of control of the Internet, see: John Gregory (2013).

[12] On precedent, see: Samuel Warren and Louis Brandeis (1890).

[13] On the right to privacy, see, among others, the following sentences of the Spanish Constitutional Court: Constitutional Tribunal (2011), Second Chamber, Sentence 173/2011, of the 7th of November, 2011; Official State Bulletin, No. 294. Wednesday 7th of December, 2011, pp. 1-15. Constitutional Tribunal (2013), First Chamber, Sentence 170/2013, of the

7th of October, 2013; Official State Bulletin, No. 267, Thursday, 7th of November, 2013, pp. 49-67.

[14] The initial regulation was the law on Protection of Personal Data of the 7th of October, 1970, passed by the State of Hessen, Germany. The law is known as the "*Hessisches Datenschutzgesetz*". Published in *Gesetz - und Verordnungsblatt für das Land Hessen* (1970), number 41, 12.10.1970, pp. 625 ss. The full text can be seen at: http://www.hessischer-landtag.de.

[15] "The United States is the only Western democracy without a comprehensive law to control abuse of personal data", Lee (2012, p. X).

[16] Sentence of the European Justice Tribunal (Main Chamber), 13th of May 2014. http://curia.europa.eu/juris/document/document.jsf?docid=152065&doclang=ES

[17] For the text approved: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679. The regulation shall apply from 25 May 2018.

[18] For the summarised text (by the Spanish Data Protection Agency, see http://www.agpd.es/portalwebAGPD/CanalDelCiudadano/derechos/principales_derchos/index-ides-idphp.php

[19] A detailed study on Internet regulation can be found in Koukiadis (2014).

[20] In this section we refer to the work of Fernando Galindo (2014).

[21] The following definition, as a guideline, is rather generic: "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)". See: http://opendefinition.org. A second, less ambiguous definition: "A piece of data or content is open if anyone is free to use, reuse, and redistribute it - subject only, at most, to the requirement to attribute and/or share-alike". See: http://opendefinition.org/, the consultation took place on the 16th of October, 2014. It is not sufficiently precise if the search for data is just undertaken with search engines such as Google.

[22] http://www.europarl.europa.eu/charter/pdf/text_en.pdf

[23] For an authoritative briefing on thesaurus evolution and concept, see: Aitchison, and Dextre Clarke (2004) and Dextre Clarke and Zeng. (2012).

[24] Others propose the use of "anonymous query" methods but anonymization can have other risks: e.g. the use of statistics, and they only generate economic benefits for the use of the search engines. See: Pàmies-Estrems, D., Castellà-Roca, J., Viejo. A. (2016).