DOI: 10.13166/WSGE//VZHR8757

Weronika Pielak-Sitek
University of Warsaw
ORCID ID: 0000-0003-2823-9083
Wojciech Sitek
Alcide De Gasperi University of Euroregional Economy in Józefów
ORCID ID: 0000-0002-1545-182X

INFORMATION RETRIEVAL AGAINST THE HUMAN RIGHT TO PRIVACY

Abstract
In the era of digital transformation, the main human right to be protected on the Internet appears to be the right to privacy. Human rights are breached not only by the governments and military forces, but also by the international private corporations. The rapid development of the Information Retrieval methods with the Machine Learning techniques and unrestrained access to personal data gives global potentates access to automatic processing of personal Big Data. In the article there are discussed the vital problems of the privacy of the humanity, the need for international regulations for this human right enforcement and the reflections over uninhibited, technical expansion without ethical boundaries.

Keywords: *privacy, human rights, digital transformation, PRISM, Information Retrieval*

Introduction

The concept of human rights protection originates from philosophic and Judeo-Christian value system. After the II World War and because of its catastrophic consequences, an urgent need for the legal and social protection of individuals has emerged. As a result, in 1948, the Universal Declaration

of Human Rights, followed by The Convention for the Protection of Human Rights and Fundamental Freedoms in 1953, has been established. The legal activities evoked the science interest, the launch of promoting and monitoring human rights in the world by organizations and foundations and incorporating these concepts into national laws.

However, the development of human rights includes only superficially the process of digital transformation and its consequences. An advance in technical possibilities of computer science with the changing world economy and the way of living has been faster than the background for the transformation in the other disciplines. The legal regulations, the solutions of the ethical dilemmas and psychology analysis follow the outdated information technology issues from the decades ago and develop slower than the digital transformation. This situation causes the ideal conditions for the unrestrained and unreflective progress in computer science.

An increase in speed, reliability, security, number of applications and availability of the Internet has caused its wide use in personal life of individuals and different sectors of the economy. A significant proportion of the tasks undertaken in the World Wide Web cover processing the text, multimedia, and other resources. Using the possibility of automatic and manual processing of the huge quantities of resources available on the Internet has become a significant source of knowledge, but also a great danger for the human rights protection. Thus, the human rights need to be reflected and adapted in the cyberspace context.

The human right, which is particularly exposed to the negative aspects of the Information Retrieval (IR), is the right to privacy. According to the basic international conventions, the right to privacy is one of the primary individual rights of the person.

In this article, it will be discussed the need for new interpretation or update of human rights legal regulations regarding cyberspace. Because of the breadth of the human rights catalogue and the computer science disciplines catalogue, only the right to privacy and the IR field will be studied. Apart from individual view on the subject, it will be discussed the impact of huge international corporations on the rights of the individuals. The descriptive methodology will be used, based on the analysis of the related literature and on the ordered regulatory observations endorsed by technical computer science findings.

In the Section 2 there will be provided the human rights background. In the Section 3 it will be provided the legal analysis of the conventions upon the right to privacy. In the Section 4 there will be described novel information retrieval methods and their strength in automatic multimedia processing, followed by the discussion of information retrieval role against the right to privacy. The last Sections 5 and 6 will be dedicated to the superior examples of usages of the IR by the international companies – the social media and Internet searchers. The article will be finished by the conclusions.

Human rights background

Most of the current technical Computer Science publications cover an invention of a new solution for the commonly addressed technical problems. They include methods, experiments and technical discussion of results and their parameters like time and memory complexity. However, the strict nature of the technical studies does not raise interdisciplinary issues of law regulations, human rights protection, and humanistic reflection over new solution. On contrary, the human rights science research publications focus rather on the classical issues from the letter of the international conventions, i.e., slavery, discrimination, arbitrary arrest, cruel treatment, or persecution. Thus, the interdisciplinary approach to the human rights in the cyberspace is only the alternative research problem, which needs to follow the technical progress.

From the legal and ethical point of view, it is stated[1] that the humanity entered the new epoch of virtual creation instead of possessing and using previously created reality. Therefore, the creators of the new virtual space should also create the ethical code. On contrary, it is also said[2] that the cyberspace is only the complex set of the technological features that enables new possibilities for the humanity, compared with the invention of a car or a bicycle. It is concluded1 that the ethical rules in the cyberspace should be adopted from classical ethics by the interdisciplinary bodies.

[1] Sitek B. (2016). Zasady etyczne stosowane w cyberprzestrzeni. W: B. Sitek, J. Knap, S. Sagan, Ł. Roman, Nowoczesne narzędzia informatyczne w przeciwdziałaniu zagrożeniom bezpieczeństwa, Józefów: Wydawnictwo WSGE, 71-84.)

[2] Ploug, T. (2009). Ethics in cyberspace (pp. 3-12). Springer Netherlands.

The social and legal analysis of the human rights in cyberspace with the mention of the need for special protection and education has been presented in Florek, 2018[3]. Basing on the same legal ground, in this article more technical threads for the IR field will be shown.

Yet it is stated[4] that the Internet, as not fully regulated, became a place for abuses of big corporations, changing the initial assumptions of World Wide Web. It is stated that the Internet is a place of great contradictions and without regulations, it will not change significantly.

As it can be observed from existing literature regarding the subject of the article, in the provided human rights articles, almost only humanistic observations are taken into consideration and discussion. Thus, in this article, basing on the humanistic observations, more interdisciplinary and more technical analysis will be provided.

Right to privacy – legal analysis

According to the Article 17 of the International Convenant on Civil and Political Rights:

1. *No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation.*
2. *Everyone has the right to the protection of the law against such interference or attacks.*

Following privacy publication of Humble[5], application and interpretation of this article has become the basis for the international conventions and the privacy treatment in many UN member states. These issues were furtherly unfolded in the General Comment No 16 on Article 17 ICCPR and regulated in continental conventions like ECHR.

According to the Article 8 of the European Convention on Human Rights:

[3] Florek, I., Eroglu, S. E. (2018). The need for protection of human rights in cyberspace. Journal of Modern Science, 2019, 27-36.

[4] Steeves, V. (2000). Privacy, Free Speech and Community: Applying Human Rights Law to Cyberspace. In Human Rights and the Internet (pp. 187-199). Palgrave Macmillan, London.

[5] Humble, K. P. (2021). International law, surveillance and the protection of privacy. The International Journal of Human Rights, 25(1), 1-25.

1. *Everyone has the right to respect for his private and family life, his home and his correspondence.*
2. *There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.*

When the Convention has been established, it did not cover the digital development and its consequences. The Convention covers the television, radio, cinema, press, postal services, telegraph, but needs to be transposed for the mobile telephone, Internet, social media, Internet searchers, digital databases and the reality of intelligent algorithms.

ECHR focuses on minimization of intervention of the state into the private areas of the individuals. However, in the times of a vital impact of digital technology on everyday life, most Internet service providers are owned by international private companies. Thus, the state in which there is no company headquarters is often stripped from the legal and political impact on the organizational, legal, and technical regulations which are established internally in the company and their enforcement.

Nevertheless, in the case of *Schrems v Data Protection Commissioner* by the Court of Justice of the European Union (CJEU), as it is observed5, giving access of the personal data of the Europeans to the non-EU country USA by Facebook, is against European treatments and the GDPR Regulation. However, the CJEU did not observe that the private communication company Facebook has the capability to automatically process the data of millions of Europeans with no strict control and regulations.

It is observed[6] that National Security Agency in the USA and the Communications Headquarters of the Government of the UK lead mass surveillance of the entire populations of international community. After the revelations of E. Snowden, it is also known that the USA PRISM program enables automatic processing of multimedia from VoIP conversations, messages, social media, and other strictly private sources. They conclude that

[6] Watt, E. (2017). The right to privacy and the future of mass surveillance. The International Journal of Human Rights, 21(7), 773-799.

there is an urgent need for the supplement of the ICCPR and ECHR and new international Intelligence Codex, which would enforce the protection of the right to privacy in the cyberspace.

However, it is also stated[7] that the mass surveillance is becoming a new Internet and influences not only the right to privacy, but also the rights of freedom of thought, conscience, religion, expression, the right to a fair trial, the freedom of assembly and association and the prohibition of discrimination[8]. They give examples of profiling based on racism and religion intolerance or claiming of predicting the human behavior by the digital assistants better than humans evaluate themselves.

To entirely understand the legal and organizational issue regarding international corporations, human right to privacy and create an interdisciplinary overview, it is needed to analyze the technical capabilities of the surveillance subjects.

Automatic multimedia processing against the right to privacy

In the ancient times, there were great needs for high computation power[9]. Egyptian priests have counted the dependencies between the stars. The Romans prepared statistical information about people and the taxes. In these times, only human clerks were doing manual computations. Moreover, the statistical information about the books, libraries and the natural language was also partial. There was not any fast possibility to count the statistics of the words in all books from the library or to recognize the writer by their style character. All the computations made on a few most popular books like Bible or Cicero pieces were extremely tedious and took many years to be prepared.

---

[7] Bernal, P. (2016). Data gathering, surveillance and human rights: recasting the debate. Journal of Cyber Policy, 1(2), 243-264.

[8] Wieczorek-Płochocka W. (2020) Pandemia a prawo do prywatności [in:] Burda E., Lazaro Guillamon C., Sitek M., State and society facing pandemic, Publisher Comenius University in Bratislava, p. 350. https://doi.org/10.13166/mng/100036

[9] Haga, E. J. (1962). History of Digital Computing Devices. School Science and Mathematics, 62(3), 197-205.

After an invention of a computer, a new perspective for a fast automatic computing has emerged. Information retrieval is the field in computer science started by Gerard Salton in his important publication[10], about vector space model. He stated that embedding word and sentence representations to vectors will make computer text processing more efficient.

Through the XX century, the methods were mainly traditional, based on text frequency analysis and heuristics. The architectures were multi-modular, complicated and had large time complexity. Moreover, often the whole corpus had to be accessible during all the training and prediction time.

In the XXI century, when the Artificial Intelligence systems are gaining popularity among the researchers, the Machine Learning systems dominated the Information Retrieval field. The Word2Vec architecture[11] consists of the neural network trying to predict the embedding vectors of neighbor context words from the central word or the central word from a few context words. After training the embedding vectors of the entire vocabulary, it is possible to indicate the most similar words or to automatically process the vectors in a multidimensional space instead of the categorical vectors, which largely increases the effectivity of text analysis.

Doc2Vec model[12] enables automatic comparison of the complete documents instead of the words, because of training the document vector along with the word vectors. Using this architecture seems a very useful method of obtaining the automatic evaluation of author's style and vocabulary and may be used for further IR tasks. TF-IDF, FastText, GloVE, ELMo and many other techniques are used for word, phrase, and document embeddings.

Another task which enables mass surveillance techniques is the development of the machine translation. Bidirectional RNN, GRU, LSTM, Attention, Transformers, BERT, Autoencoders, GANs[13] and many methods have emerged

[10] Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11), 613-620.

[11] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[12] Le, Q., & Mikolov, T. (2014, June). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196). PMLR.

[13] Koehn, P. (2017). Neural machine translation. *arXiv preprint arXiv:1709.07809*.

in the last few years and enhance the effectiveness of translation of articles, private messages, voice records and searcher result tagging[14].

In general, Deep Neural Networks in the different configurations and variants make many fuzzy, indeterministic, human-skilled tasks possible to process by computers. Regression and classification tasks may find different applications also useful for surveillance like predicting future time sequence of socio-economical occurrences, converting speech to text, assigning keywords to texts of different lengths, describing images by texts, recognizing people from the images, or estimating the danger of terrorism of individuals basing on the surveillance data.

The advanced algorithms and structures enable mass data aggregation about private life, personality features, social environment, political views, and many other areas of private and family life basing on user activity on the Internet. Average web user is not aware that almost all the content placed in the social media is profiled by the algorithms for automatic processing of text, audio, images, and video. They can be categorized to the different groups by views, preferences, and sensitive information about the personality. Then, this information may be used without real restrictions by the service provider for other purposes.

Moreover, all the Internet services are exposed to the cyber-attacks, which may lead to uncontrolled data breaches on a scale of the hundreds of millions of accounts[15]. The data breaches usually cover private and sensitive personal data, like address, credit card numbers, or social platform messages and posts.

[14] Nayak, P. (2019, October 25). *Understanding searches better than ever before*. Google. https://blog.google/products/search/search-language-understanding-bert/

[15] Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The guardian*, *17*, 22.

Privacy in the social media

Social media and Internet forums constitute previously unknown case, where each user of a service is at the same time a co-author of the emerging content.

The role of service provider is constrained by the essential actions:

- providing the technical background as maintaining the physical layer in connection with application layer for the users from many countries in the world, as well as storing current and historical contents of the medium.
- automatic matching and proposing to the users' contents profiled to them by their history, provided information and their behavior in the medium. It is usually obtained by the intelligent algorithms.
- constituting the regulations of the acceptable behavior of the users (i.e. morality, prohibition of vulgarity, hate, criminal actions) and enforcing them (i.e. by notifying the state services, blocking access to the account for a given period or deleting unacceptable content). Moreover, often these actions of the corporations are undertaken automatically.

The novel important aspect regarding social media is to move the creative responsibility from the professional subjects, like press and television companies, to the completely unprofessional subjects – each person with Internet access, including children.

The consequences of becoming the creator are the increase in creativity for promoting own point of view of the world, but also for self-promotion. It can be observed that often for the individuals themselves, the overarching objective over respect to their own private, family life and other moral values becomes their popularity. As a result of the peer pressure, the private information becomes public on the Internet and the individuals resign from their privacy.

The resignation from the usage of the right to privacy seems voluntary. However, the common pressure of the society and the increasing convenience of using social media with no alternative may decrease the level of voluntariness with no legal actions against international conventions.

Moreover, the aspiration of popularity in social media is very strong among children and is amplified by the advertisements and the culture of consumption. Children, once addicted to the social media, often remain overusing this communicating medium for the rest of their lives.

Searcher privacy protection

The technical possibilities of the Internet during last decades are developing faster than law regulations regarding this subject. Additionally, decentralized specifics of the Internet make enforcing the regulations by the national and international law much harder[16]. With the popularization of the Internet, a few propositions of approaching virtual space legally have emerged – as a *stricte* international space; where the responsibility depends on the citizenship of a user; finally as a space with no law regulations.

A great danger for enforcing the right to privacy is impossibility of total enforcement of the right to be forgotten by the international searcher and the Artificial Intelligence architectures[17]. It is concerning that it is possible that even if the information is evaporating from the social media, searchers and websites, it only becomes archived, invisible for the user, but it is still possible to access by the administrators of the system and the intelligence services.

[16] Gałka, P., & Ciach, S. (2019). Blockchain a prawo: mapa drogowa najważniejszych zagadnień. *Krytyka Prawa*, *11*(2), 40-51.

[17] Villaronga, E. F., Kieseberg, P., & Li, T. (2018). Humans forget, machines remember: Artificial intelligence and the right to be forgotten. *Computer Law & Security Review*, *34*(2), 304-313.

Conclusions

Automatic text processing is a glorious conquest of the modern civilization and has many useful applications, but their usage in mass surveillance on the Internet poses a dominant concern about the respect for the right to privacy of the world population. Intelligence services' global programs revealed by E. Snowden in 2012, like PRISM program with the participation of Microsoft, Apple and Google, seem frightening for non-USA governments and human rights organizations.

As a conclusion, it seems vital for the human rights protection to provide the strict international law regulations regarding the right to privacy, prepared by the interdisciplinary bodies of computer scientists and technology-aware lawyers. Not only should the law regulations be introduced, but also, they should be strictly enforced on the great international corporations like Facebook and Google.

Finally, the research over automatic text processing and other technical issues should include a deep ethical and law discussion. Moreover, lawyers and ethicists should be educated in technical elements of Computer Science to increase their awareness and the quality of their work upon technical issues.

References

Bernal, P. (2016). *Data gathering, surveillance and human rights: recasting the debate.* Journal of Cyber Policy, 1(2), 243-264.

Cadwalladr, C., Graham-Harrison, E. *(2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach.* The guardian, 17, 22.

Florek, I., Eroglu, S. *E.* (2019). *The need for protection of human rights in cyberspace.* Journal of Modern Science, 27-36.

Gałka, P., Ciach, S. *(2019). Blockchain a prawo: mapa drogowa najważniejszych zagadnień.* Krytyka Prawa, 11(2), 40-51.

Haga, E. J. *(1962).* History of Digital Computing Devices. School Science and Mathematics, 62(3). 197-205.

Humble, K. P. *(2021).* International law, surveillance and the protection of privacy. The International Journal of Human Rights, 25(1), 1-25.

Koehn, P. (2017). *Neural machine translation.* arXiv preprint arXiv:1709.07809.

Le, Q., Mikolov, T. *(2014). Distributed representations of sentences and documents.* In International conference on machine learning, 1188-1196). PMLR.

Mikolov, T., Chen, K., *Corrado, G.,* Dean, J. (2013). E*fficient estimation of word representations in vector space.* arXiv preprint arXiv:1301.3781.

Nayak, P. (2019). *Understanding searches better than ever before.* Retrieved from: 2019, October 25. Google. https://blog.google/products/search/search-language-understanding-bert/

Ploug, T. (2009). *Ethics in cyberspace, 3-12.* Springer Netherlands.

Salton, G., Wong, A., *Yang, C.* S. (1975). *A vector space model for automatic indexing.* Communications of the ACM, 18(11), 613-620.

Sitek, B. (2016). *Zasady etyczne stosowane w cyberprzestrzeni.* W: B. Sitek, J. Knap, S. Sagan, Ł. Roman. Nowoczesne narzędzia informatyczne w przeciwdziałaniu zagrożeniom bezpieczeństwa. Józefów: Wydawnictwo WSGE, 71-84.

Steeves, V. (2000). *Privacy, Free Speech and Community: Applying Human Rights Law to Cyberspace.* In Human Rights and the Internet, 187-199. Palgrave Macmillan. London.

Villaronga, E. F., *Kieseberg, P.,* Li, T. (2018). *Humans forget, machines remember: Artificial intelligence and the right to be forgotten*. Computer Law & Security Review, 34(2), 304-313.

Watt, E. (2017). *The right to privacy and the future of mass surveillance.* The International Journal of Human Rights, 21(7), 773-799.

SOCIETY UNDER THE INFLUENCE OF TECHNOLOGY