NATIONAL UNIVERSITY
 OF PUBLIC SERVICE
Doctoral Council

**JÁNOS PÁL VADÁSZ**

*- The methods and application possibilities of semantic search in the defense, in public service and in the economic sphere-*

Author's summary and official reviews of the PhD dissertation

Budapest
2018

**NATIONAL UNIVERSITY OF PUBLIC SERVICE**

JÁNOS PÁL VADÁSZ

*- The methods and application possibilities of semantic search in the defense, in public service and in the economic sphere-*

Author's summary and official reviews of the PhD dissertation

Supervisor:

Prof. Dr. Sándor Munk col. ret. (DSc)
professor

Budapest
2018

# THE DESCRIPTION OF THE SCIENTIFIC PROBLEM

Though information search has been playing a major role in the fields appearing in my dissertation subject (defense, public service and economy), scientific research has focused little on the interdisciplinary areas of research subjects, even less in the Hungarian local context. There are many publications on the mathematical and informatics aspects of my subject, while we can read few on the usability, feasibility, and legal ecological aspects of information search. It seems that this area is too practical for theoretical scientists, while too theoretical for practitioners.

As a manager of a company providing services in the field, I have always been interested in information search and specifically in semantic search. My experience has shown that there are many issues in the micro- and macro-social environment of information search, that the Hungarian scientific literature hardly covers, and even the international one does not favour. These include, among other things, the following.

As information and its effective search for has become an independent resource in the area of the economy and public administration, information search has also turned out to be an investment target on its own. However, the information search as an investment area is practically unknown in Hungarian industrial and administrative circles. Senior management is often reluctant to take investment decisions due to the lack of knowledge required, and this hinders technological renewal and efficiency improvement. Even the simplest model based on time saving shows startling results. If we add to this the intangible characteristics that can be found by indirect measurements only, the decision-makers are given complete arsenal to make their investment decisions. However, their scientific examination is incomplete, essentially lacking.

Another unexplored but also an important aspect of the micro-environment is the human and security side of the information search, the detection of barriers that prevent the absorption of an information search application by the working organisations. The exploration and systematisation of these criteria based on domestic industrial experience and international scientific literature can also be identified as a scientific problem.

Finally, I consider it an important scientific problem to systematise the framework of the Hungarian legal system from the viewpoint of information search, to identify gaps and shortcomings that prevent progress of technological progress. As the background to domestic

research, I consider the international outlook to be a research task to highlight the direction of the development of legislation in the EU and in the Anglo-Saxony countries.

During my nearly 30 years of professional history in Hungary, I tried to build a bridge between scientific theory and industrial practice. Combined the results of my research with decades of industrial experience, I started the endeavour of writing this dissertation.

## RESEARCH OBJECTIVES

The basic research objective of my dissertation is to analyse the effective and efficient conditions of information search in the defence sphere, public administration and economic life. As part of this:

- exploring the priority areas of the application of information search in the defence sphere, public administration and economic life;
- exploring and evaluating the Hungarian legal environment of search of information, identifying the legal norms hindering information search and identifying their impact;
- identifying and establishing methods for the accountability of law enforcement and national security organisations, that increase the confidence of citizens on the one hand and the effectiveness of the operational environment of these organisations on the other;
- the development of a framework and a model for the efficiency and effectivity of information search;
- identifying obstacles to the introduction of information search as a knowledge management tool and suggesting measurement to counter them.

## RESEARCH METHODS

For my dissertation I used research methods as follows:

- wide range literature overview, studying and processing the relevant international and national literature, legal and other relevant documents;
- applying secondary analysis, I analysed and processed previous research results;
- I analysed and processed the results of my professional experience, I carried out empirical research;
- I consulted leading professionals, high level managers in my research field;
- based on comparative analysis carried out on the gathered and processed information I made abstractions, conclusions and generalisations.

## RESEARCH HYPOTHESES

My research hypotheses relating to the subject of the dissertation and the elaboration of the scientific problem:

- the concept of the open source intelligence of the defence and the business sector is inaccurate and ambiguous (new scientific result 1 and 3);
- the use of semantic search systems significantly increases the effectiveness of the daily work (new scientific result 7);
- international trends show that the legal environment of the information search is undergoing a dramatic transformation (new scientific result 5);
- the way the Hungarian legal system handles the information search is fragmented, incomplete and anachronistic, the law enforcement organisations, national security services are unnecessarily hindered, and, beyond this, it constitutes a threat to the effectivity of counter-terrorism activities and efforts against organised crime (new scientific results 2 and 4);
- work organisation and communication are as important as the technology itself in the effective functioning of an information search system (new scientific result 6).

## CONCISE DESCRIPTION OF THE RESEARCH CARRIED OUT BY CHAPTER

The doctoral dissertation consists of four chapters.

In the first chapter, I summarised the theoretical knowledge base of information search, the basic concepts of the meaning-based semantic search; the mathematical and IT foundations of information search, the methods used, and finally the linguistic background of information search, the key issues of natural language processing.

In the second chapter, I presented and evaluated the three fields of application closely related to information search, its technological solutions: the sentiment analysis, the federated search and the fusion centres.

In the third chapter, I examined a primary application of information search in the three fields (defence, public administration, economic life). In each of these fields of application, I defined the most important features of the information search, justified the need for its application, its basic uses, proposed the possibilities for improvements, and projected developments in the future.

In the fourth chapter, I examined three basic external and internal environmental conditions for the effective and efficient organisational application of information search - the legal environment, human and security conditions within the organisation and the evaluation of the information search. Finally, I summarised related proposals.

## SUMMARISED CONCLUSIONS

At the beginning of the XXI. century, information technology and its use are undergoing paradigm shift. Almost all new data are in electronic form, their storage and communication functions globally and cost-effectively. At the dawn of the IV. industrial revolution, one of the key challenges is to filter out the user's relevant information, and to do this quickly, efficiently, preferably with as little human effort as possible. My research aims to explore the applicability of technologies designed to address these challenges in the area of national security, law enforcement, public administration and business. My aim is to explore the most advanced technologies of information search applications, including the computer linguistics, mathematics and IT background, and then present some of the most interesting technologies from the user perspective. Following this, I will approach the subject from a user perspective national security and law enforcement, public administration, law, and business. Finally, after the sectoral approach, I examine the information search from the organisational perspective in terms of legal, economic and organisational security.

In the first chapter of the dissertation, I cover the theoretical background of information search, with the view to the arsenal of knowledge representation, computer linguistics, mathematics and IT, up to the disruptive effects of the latest deep learning methods. I **pointed out that the searchability of the exponentially increasing text volume, of which is 80-85 %,** is **an indispensable technological challenge**. Looking through the tools of knowledge representation**, I came to the conclusion** that although technologies using classical, morphological tools and controlled dictionaries are marginalised in the context of modern statistical and deep-learning techniques for online searches, **the use of taxonomies and ontologies generated by the humans is still a long way to play a greater degree for punctuality and recall in case of enterprise content search systems**.

In the second chapter, I demonstrated the applicability of information search through three basic technologies. The three **technologies** are different in character, but they're in common in everyday life, which **I illustrated in** all three cases**.** In the federated search section, I classified the types of technology and then presented the applicability in a case study to detect illegal drug

trafficking, suggesting that the method can also be used for the collection of internet advertisements of smugglers, etc. The technology was presented at scientific conferences.

In Chapter Three, I examined information search from the perspective of users. These are the defence sector, public administration and economic life. In the analysis of OSINT, I **have reinterpreted the concept** of **open source intelligence** based on the modern Anglo-Saxon literature. In my interpretation, the essence of OSINT is not necessarily a matter of legality, but that of the public availability. I made **a sharp line** between **active and passive interventions**. I am aware that this approach is not shared with all schools, but I say that anyone who monitors and exercises the development of modern technologies is sensing the shift of the centre of gravity.

I also examined in the third chapter shortcomings in the Hungarian legal environment in relation to information search for public data. **I have revealed the legal shortcomings** that will cause EU directives not fully transposed in Hungarian law. I **pointed** out **the lack of execution instructions** and the non-enforcement of the law. I have **developed a recommendation** for **the necessary changes**.

I **redefined the concept** of **economic intelligence** and **competitive intelligence**. I have clarified conceptual disturbances in the former two, as well as in the concepts of industrial espionage and business intelligence. I have organized the technologies for information search in the field of economic intelligence.

In the fourth chapter, I examined the **Hungarian legal environment** in terms of **information search**. The study has been carried out following four concepts, so far only superficially covered by literature: the purpose limitation, prohibition of interconnectivity of databases, limitation of storage of data files over time and freedom of profiling. **I have demonstrated that the Hungarian legal system** is **in significantly backward compared to modern technological expectations**. I have revealed that, in the absence of radical change, national security services and law enforcement bodies will not be able to carry out their duties in a short time if the Hungarian legal system does not follow international trends and does not meet the challenges of technological development. **I have made recommendations** for **the necessary changes**.

After home exploration of the legal environment, I have been subjecting a thorough analysis to the so-called Anderson report, which examines the situation in the United Kingdom. The most important **statement is that bulk search is a lesser violation of personality rights, rather**

**than a targeted search**, because, though it looks at many of them superficially, it only penetrates the private sphere of those who fall under strong suspicion as a result of the mass search, so privacy of those who have nothing to do with the case is not violated.

Finally, on the basis of international scientific literature**, I have collected the methods to improve the accountability of national security services and law enforcement bodies**, so that they are limited in their technical capabilities while respecting personal freedoms. I **have proposed changes** in the **Hungarian legal system**.

I consider it a new result that I have discovered and organised factors that are barriers to the introduction of a search system and the adoption by an organisation. I found **that the main obstacle to success is not the limitations of technology, but the human factor**. **On the basis of my long industrial experience, I have proposed to combat the obstacle for each factor.**

**I have presented a model** for assessing the economics of a search systems **as a model for calculating the return on investment**, taking into account the productive use of saved professional time. Because in a knowledge-based organisation, like a legal office, court, prosecutor, market research institute, etc. the working time of highly qualified staff is the largest cost factor, the model presented gives a good approach to ROI.

## NEW SCIENTIFIC RESULTS

1. I reinterpreted the concept of open source search. I have noted that open source search does not have a definition in the Hungarian legal system. I have proposed a replacement.

2. I have revealed the methods that hinder the search for public data under the relevant EU directives and Hungarian legislation. I developed a recommendation for changes.

3. I reinterpreted the concept of economic intelligence, and I put in this context the concepts of business intelligence and industrial espionage.

4. I have revealed the Hungarian legal environment for finding information on the basis of the evaluation criteria I have developed, taking into account the Hungarian and international literature, and I have noted that the Hungarian legal environment does not follow the requirements of technological change and I have developed recommendations for changes.

5. I have examined the accountability of national security services and law enforcement bodies on the basis of international and national literature, and recommended methods

which increase the trust of citizens on the one hand and the effectiveness of of the services on the other.

6. I examined the absorption barriers to the information search as a method of knowledge management in any organisation and made suggestions to remove these obstacles.

7. I developed a case study to measure the efficiency of a legal search system based on interviews and own calculations. By using them, I drew conclusions and made recommendations for the technological renewal of the Hungarian judicial system.

## RECOMMENDATIONS

Based on the above result I make the following recommendations.

- The text or parts of the dissertation are suitable for using as teaching material or as a university note. There was no mention of information about OSIN at all in 2006, and OSIN was last mentioned in 2006, which, in addition to its excellence, requires a new approach to this discipline.

- The economic model developed can be used for the calculation of returns for investments, development of information search systems.

- The methods recommended in the human and security chapter can be useful for system analysts, evaluators and project managers on a daily basis.

- The methods recommended in the human and security chapter can be useful as education material for senior management.

- I propose the amendment of law 2015. XCVI. in two directions. On the one hand, it should be mandatory to publish the public interest data in a machine-readable form in a timely manner. On the other hand, they sanction non-compliance with this addendum.

## PUBLICATION LIST OF THE CANDIDATE RELEVANT TO THE DISSERTATION'S TOPIC

*Peer-reviewed journal articles:*

Vadász Pál, Séllei Márton: Az információkeresés magyar jogi környezete, HADTUDOMÁNY: A MAGYAR HADTUDOMÁNYI TÁRSASÁG FOLYÓIRATA 27:(1-2) pp. 178-191. (2017)

Vadász Pál: Információkeresés a nyílt forrású hírszerzésben, FELDERÍTŐ SZEMLE XIV:(1) pp. 81–100. (2015)

Vadász Pál: Semantic technologies in sentiment analysis, BOLYAI SZEMLE XIV:(4) pp. 42–51. (2015)

Vadász Pál: INFORMÁCIÓKERESÉS A GAZDASÁGI HÍRSZERZÉSBEN, HADMÉRNÖK IX.:(2.) pp. 343–357. (2014)

Vadász Pál: A metakeresés alkalmazása a bűnüldözés és felderítés világában, NEMZETBIZTONSÁGI SZEMLE (ONLINE) II:(2) pp. 58–71. (2014)

Vadász Pál: Case study for measuring the feasibility of a semantic search system, HADMÉRNÖK VII:(2) pp. 405–415. (2012)

Vadász Pál: Egy nyílt forrásokra épített szemantikus keresőrendszer bemutatása, HADMÉRNÖK VII:(2) pp. 351–359. (2012)

***Peer-reviewed publications in foreign published books:***

Vadász P, Benczúr A, Füzesi G, Munk S: Identifying Illegal Cartel Activities from Open Sources In: Akhgar B, Bayerl P S, Sampson F, edited.: Open Source Intelligence Investigation: From Strategy to Implementation. 304 p. Cham (CH): Springer, 2016. pp. 251–273., (Advanced Sciences and Technologies for Security Applications) ISBN: 978-3-319-47670-4)

***Peer-reviewed professional's conference presentation published in conference proceedings:***

Vadász Pál

A Case Study on Finding Fraudulent Practices in the Public Procurement Process Using Text-Mining Methods from Open Internet Sources in: Alexander Balthasar, Blaž Golob, Hendrik Hansen, Robert Müller-Török, András Nemeslaki, Johannes Pichler, Alexander Prosser szerk.: Central and Eastern European eIDem and eIGov Days 2016: Multi-Level (e)Governance: Is ICT a means to enhance transparency and democracy?, 607 p., Konferencia helye, ideje: Budapest, Magyarország, 2016.05.12-2016.05.13. Wien: Austrian Computer Society, 2016. pp. 471–480. ISBN 978-3-903035-11-9

## SCIENTIFIC-PROFESSIONAL BIOGRAPHY OF THE CANDIDATE

**Name:**        János Pál Vadász
**Telephone:**        + 36 30 93 11 00 3

**E-mail:**          vadasz.pal@montana.hu

**Education:**

| | |
|---|---|
| 2012 – | National University of Public Service, Doctoral Scholl of Military Engineering |
| 1975 – 1980 | Eötvös Loránd University, Faculty of Natural Sciences, M.Sc. in mathematics and physics, – secondary school teacher grade |
| 1970 – 1974 | Eötvös József Secondary School, Budapest |

**Languages:**

| language | Type of exam | Examination Center |
|---|---|---|
| English | advanced C2 complex | University of Cambridge[1] |
| English | intermediate B2 complex | ELTE ITK |
| German | advanced C1 complex | Goethe-Institute |

**Professional courses:**

| | |
|---|---|
| 1991 – 1994 | Management courses at INSEAD, Fontainebleau |

**Professional experience:**

| | |
|---|---|
| 1980 – 1985 | IBM Australia |
| 1986 – 1989 | R+S Computer-Consulting GmbH München. |
| 1989 – | Montana Group |

---

[1] It was NOT certified by NYAK/Ministry of Education, because the date of the exam was before 2000.