Factographic information retrieval for semiconductor physics, micro- and nanosystems

S D Kulik
National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashirskoe shosse 31, Moscow 115409, Russian Federation

Corresponding author’s e-mail address: sedmik@mail.ru

Abstract. This paper proposes a factographic information retrieval for semiconductor physics, as well as micro- and nanosystems. This factographic information retrieval includes a pattern recognition algorithm and factographic database. The first factographic database includes special keywords from semiconductor physics and micro- and nanosystems. The second factographic database includes document descriptions in the form of Searching Documents Patterns for the fields of semiconductor physics and microelectronics. An analytical model of the factographic information retrieval system is developed. This model is presented by an effectiveness indicator: the average length of the recommendatory list used for semiconductor physics and micro- and nanosystems for space application.

1. Introduction
The Effective Factographic Information Retrieval System is very important for scientists and researchers in semiconductor physics and for space application of micro- and nanosystems [1, 2, 3] to retrieve factographic information with keywords. Next, without loss of generality, we may assume that some keywords are in papers, for example, scientific papers [4-25]. In our research, we used effective information technology.

Professional competencies [18, 19, 27, 28] are very important for scientists and researchers in various areas - for instance, in microelectronics. Factographic information retrieval also improves professional competency at times.

Thus, the paper is organized as follows. A brief introduction and short review of the research field are introduced in Section 1. The research methods are introduced in Section 2. Section 3 contains some examples and keywords (special features of scientific texts) of documents of semiconductor physics and micro- and nanosystem space application. The section contains a brief description of the factographic information retrieval system. Results, discussion and future works are introduced in Section 4. The section contains analytical formulas to evaluate the effectiveness of factographic information retrieval for semiconductor physics and microelectronics. The key defining indicator of factographic information retrieval in the factographic retrieval system is introduced in Section 4. This section presents the results obtained from analytical expressions (analytical models) and brief discussion of the results. Finally, the main results are outlined in Section 5. Some aspects are not discussed in detail because of the article’s limited volume.
Many problems were solved and were studied in detail. However, some important problems remain unsolved, for example:

- development of an effective FIRS for application in semiconductor physics and space micro- and nanosystems;
- recognition of new experimental and theoretical results automatically from scientific texts;
- recognition of new scientific ideas from the database for semiconductor physics and micro- and nanosystems for space application;
- recognition of psychological characteristics of a person by handwriting.

Artificial neural networks (ANNs) [11] have been used for Factographic Information Retrieval System (FIRS).

There are important scientific areas of applications for FIRS, artificial neural networks [11], information technology and artificial intelligence methods:

- physics and chemistry [24];
- electric insulators (glass and ceramics) [25];
- semiconductor physics, micro- and nanosystems;
- medicine [22, 23];
- police, criminology [20] and forensic experts;
- retrieval system;
- ill-posed approximation problems [12, 13];
- Data Analysis [14];
- housing market [21];
- Gaze-Based Control [15];
- Eye-Brain-Computer Interface [16];
- scientific recommender systems [17];
- knowledge management system [19];
- Big Data, Fast Data [26];
- Information Security [27, 28];
- competences forming [18];

and many other areas.

The effective factographic information retrieval problem has been investigated in the paper for semiconductor physics and micro- and nanosystems for space application.

It is necessary to give a short review of the research field. Some important problems of semiconductor physics and micro- and nanosystems for space application were considered in the indicated papers [1, 2, 3, 4, 5, 6]. Paper [1] explores a conversion model of radiation-induced interface-trap buildup. Paper [2] deals with conventional silicon bipolar transistors. Paper [3] deals with the formation of dielectric silicon compounds. Paper [4] deals with an analytical model of plasma-chemical etching in a planar reactor. Reactive ion etching of silicon using a low-power plasma etcher is introduced in [5]. Sensitive elements are introduced in [6]. Numerous papers and monographs are concerned with very different important topics, for example, neural networks theory in [12, 13], data analysis in [14], Gaze-Based Control in [15], Eye-Brain-Computer Interface in [16], recommender systems in [17], software simulators for the evaluation of socio-personal competence in [18] and cognitive competence of graduates in [19], artificial intelligence methods in [20] and Big Data in [26]. Monograph [11] is concerned with the problem of artificial neural networks. The problem of factographic information retrieval for semiconductor physics and microelectronics is very important. Despite a vast amount of literature, this problem has never been studied from a new viewpoint.
2. Methods
The research was conducted on the basis of the methods of semiconductor physics, micro- and nanosystems [1, 3, 5, 7], information retrieval, operations research, Big Data [26], artificial intelligence methods [14, 20], neural networks [11, 12, 22, 23], mathematical modeling [13, 21], competence forming [18, 19, 27, 28] and pattern recognition.

3. Factographic information retrieval system
The factographic information retrieval system includes FDB I and FDB II. FDB I includes keywords of documents of semiconductor physics and micro- and nanosystems for space application. FDB II includes factographic information of documents.

Special features of scientific texts, for example, keywords, are very important for the factographic retrieval system. Some examples of these keywords (special features of scientific texts) are shown in Tables 1, 2, 3 and 4.

Table 1. Keywords for the factographic database from [1].

| Special keywords (from paper)                   | Identification code |
|------------------------------------------------|---------------------|
| bipolar device                                 | B273d1078           |
| radiation degradation                          | R105d978            |
| averaged dose rate                             | A580d57R8           |
| silicon forbidden gap                          | S700f1G02           |
| thermoactivated tunneling mechanism            | T400t2M01           |
| space application                              | S007a567            |
| radiation-induced degradation                  | R105i04D4           |
| transition time of the electron                | T207t30E07          |
| hydrogen-electron concept                      | H005e07C1           |
| hydrogen-electron mechanism                    | H005e07M01          |
| averaged temperature                           | A058t800            |
| conversion model                               | C057m054            |

Table 2. Keywords for the factographic database from [2].

| Special keywords (from paper)                   | Identification code |
|------------------------------------------------|---------------------|
| bipolar transistor                              | B273t105            |
| SiGe-transistor                                 | SG10t105            |
| silicon transistor                              | S700t105            |
| bipolar device                                  | B273d17             |
| SiGe device                                     | SG10 d17            |
| electron-hole                                   | E007h03             |
| whole oxide layer                               | W702o5L04           |
| emitter junction                                | E201j310            |
We can assume that the \( \Omega \) is a set of special keywords (see Table 1 and Table 2) and \( Z_j \) is the element of \( \Omega \), i.e. \( Z_j \in \Omega \) and

\[
\Omega = \{ Z_1, Z_2, Z_3, ..., Z_j, ..., Z_m \}.
\]

**Table 3.** Keywords for the factographic database from [3, 4, 5, 6, 7, 8, 9, 10].

| Special keywords (from title) | Identification code |
|------------------------------|---------------------|
| dielectric silicon compounds | D148s60C98          |
| plasma-chemical etching in planar reactor | P885c9e16p08R203 |
| low-power plasma etcher      | L015p84p77E55       |
| dielectric membrane structures | D148m888S500       |
| reactive magnetron sputtering | R445m22S054        |
| high temperature annealing   | H021t800A404       |
| dielectric membrane          | D148m888            |

**Table 4.** Keywords for the factographic database from [19, 20, 21, 22, 23, 24, 25].

| Special keywords (from title) | Identification code |
|------------------------------|---------------------|
| micro-arc oxidation          | M61a43091           |
| electric insulators          | E002i28             |
| neural networks              | N622n15             |
| new knowledge                | N222k34             |
| knowledge management system  | K34m34S505          |
| intelligence methods         | I111m324            |
| mass appraisal               | M0017a008           |

Without loss of generality, we can assume that the \( \hat{\Omega} \) is set of special keywords from title only (see Table 3 and Table 4) and \( \hat{Z}_j \) is the element of \( \hat{\Omega} \), i.e. \( \hat{Z}_j \in \hat{\Omega} \) and

\[
\hat{\Omega} = \{ \hat{Z}_1, \hat{Z}_2, \hat{Z}_3, ..., \hat{Z}_j, ..., \hat{Z}_m \}.
\]

The factographic information retrieval system (see Figure 1) for semiconductor physics, micro- and nanosystems (i.e. MN-systems) consists of different blocks.
Figure 1. Factographic information retrieval system for semiconductor physics (MN-systems).

The factographic information retrieval system includes seven blocks: the block 1 of document indexing; the block 2 of recognition (neural networks); the block 3 of Factographic database I (Special keywords: $\Omega$ and $\hat{\Omega}$); the block 4 of Factographic database II (factographic information) which includes document descriptions in the form of Searching Documents Patterns (SDP); the block 5 of RL formation (retrieval block); the block 6 archive of documents; the block 7 of Recommendatory List (RL) processing. It is supposed that there are $N$ documents which are stored in the archive and every enquiry has Searching Enquiry Pattern (SEP). Factographic database II has $N$ SDPs.

4. Results, discussion and future works
Experiments and analytical results for estimation of error recognition and evaluation of the effectiveness of retrieval are very important for us.

It is supposed that $L_r$ is average length of the RL. Analytical formula was obtained to evaluate the factographic information retrieval effectiveness with the help of indicator (see (1) and (2)):

$$
L_r \approx F\left(N, \bar{L}, P_2\right).
$$

It is supposed that:

- $P_1$ – probability of the correct comparison of two identical documents based on their descriptions (determines the target mission probability);
- $P_2$ – probability of the correct comparison of two non-identical documents based on their descriptions (determines the false alarm probability);
\[
E_{i}^{N} = \binom{N}{i} = \frac{N!}{i!(N-i)!},
\]
were \(N \gg i\).

The following analytical expressions were obtained to evaluate the effectiveness:
\[
F(N, L, P_2) = \sum_{i=0}^{L-1} E_i^N \psi_i + L \sum_{i=L}^{N} E_i^N \psi_i, \quad \psi_i = P_2^{N-i}(1-P_2)^i.
\]

As a result of researches (1) and (2), it was set that \(L\) in (1) is changed during the changes of \(L, N,\) and \(P_2\). A small part of this research for different \(L\) we can see in Table 5.

**Table 5.** Example of the effectiveness, if \(N\)=50000 and \(P_2\)=0.995

| \(L\) | \(\bar{L}\) |
|------|------|
| 5    | 5    |
| 7    | 7    |
| 15   | 15   |
| 50   | 50   |
| 140  | 140  |
| 240  | 237  |
| 245  | 241  |
| 850  | 250  |
| \(2 \cdot 10^3\) | 250 |
| \(7 \cdot 10^3\) | 250 |

According to Table 5, if \(P_2\)=0.995, \(N\)=50000 (\(N\) – the number of SDP) and \(L\)=2000 (\(L\) – maximum length of the RL), the length of the RL of the FIRS is \(\bar{L} \approx 250\).

Future steps of our work include getting the answers to the following important questions:

1. How do parameters of the algorithm of pattern recognition influence information retrieval for semiconductor physics and micro- and nanosystems for space application?
2. Could one recognize new experimental and theoretical results automatically?
3. Which structure of neural network is most suitable for recognizing scientific texts from semiconductor physics and micro- and nanosystems space application?
4. Could one recognize thought from scientific texts for semiconductor physics?
5. Could one recognize new scientific ideas from database for semiconductor physics and micro- and nanosystems space application?
6. How do parameters of neural network model influence error discrimination (classification) scientific texts from semiconductor physics, micro- and nanosystems for space application?
7. Could one extract features of special of scientific texts automatically?
8. What activation function is most suitable for recognizing scientific texts for semiconductor physics and microsystem electronics?

The special neural network algorithm will be the subject of our future work in which we are aiming to improve factographic retrieval tools for superconducting microelectronics, solid-state microelectronics and microsystem electronics. That is why our plans are:
• Developing the special neural network algorithm for recognizing special keywords for superconducting microelectronics, solid-state microelectronics and microsystem electronics.
• Experiment with structures of special neural networks.
• Developing the special neural network algorithm for recognizing the scientific theoretical facts from scientific texts for other space applications.

5. Conclusion
Thus, as a result of research, the analytical formula (2) allowing the evaluation of the RL average length was developed. Set of keywords were explored for microelectronics. Factographic database I (special keywords: set $\Omega$ and $\hat{\Omega}$) was created for semiconductor physics and MN-systems.

Properties of the important indicator of effectiveness – the RL average length for the human operator were explored. Necessary software, allowing the evaluation of the retrieval’s effectiveness, was created for the developer of the Factographic Information Retrieval System for semiconductor physics and micro- and nanosystems for space application. The knowledge of Factographic information retrieval can be used as illustrative material in teaching modern information technology for microelectronics in the university.

We believe that our results are useful for scientists and researchers in the fields of semiconductor physics and micro- and nanosystems for space application to retrieve factographic information.

Acknowledgements
This work was supported by Competitiveness Growth Program of the Federal Autonomous Educational Institution of Higher Education National Research Nuclear University MEPhI (Moscow Engineering Physics Institute).

We would like to express our gratitude to students of NRNU MEPHI for their valuable discussion and suggestions.

References
[1] Pershenkov V S, Sogoyan A V and Telets V A 2016 IOP Conf. Series: Materials Science and Engineering 151 012001
[2] Bakerenkov A S, Felitsyn V A and Rodin A S 2016 IOP Conf. Series: Materials Science and Engineering 151 012013
[3] Veselov D S and Voronov Yu A 2016 Journal of Physics: Conference Series. 747 Art. No. 012022
[4] Veselov D S, Bakun A D, Voronov Yu A, Kireev V Yu and Vasileva O V 2016 Journal of Physics: Conference Series 748 012015
[5] Veselov D S, Bakun A D and Voronov 2016 Journal of Physics: Conference Series. 748 012017
[6] Veselov D S 2015 Physics Procedia 72 pp 499-502
[7] Veselov D S, Voronov Yu A and Vanyukhin K D 2015 Physics Procedia 72 pp 494-498
[8] Veselov D S and Voronov Yu A 2015 Physics Procedia 71 pp 423-427
[9] Veselov D S 2015 Applied Mechanics and Materials 799-800 pp 994-997
[10] Veselov D S and Voronov Yu 2015 Applied Mechanics and Materials 799-800 pp 919-922
[11] Haykin S 1999 Neural Networks. A Comprehensive Foundation (Second Edition, Pearson Education, Inc.)
[12] Kruglov I A, Mishulina O A and Bakirov M B 2012 Neurocomputing 96 pp 74-82
[13] Kruglov I A and Mishulina O A 2013 Journal of Computer and Systems Sciences International 52 4 pp 503–518
[14] Skiteva L, Trofimov A, Ushakov V, Malakhov D and Velichkovsky B 2016 Biologically Inspired Cognitive Architectures for Young Scientists Proceedings of the First International Early Research Career Enhancement School 449 pp 135-140
[15] Shishkin S, Nuzhdin Y, Svirin E, Trofimov A, Fedorova A, Kozyrskii B and Velichkovsky B 2016 Frontiers in Neuroscience 10 pp 1-20
[16] Nuzhdin Y O, Shishkin S L, Fedorova A A, Trofimov A G, Svirin E P, Kozyrskiy B L, Medyntsev A A, Dubynin I A and Velichkovsky B M Proceedings of the 2017 ACM Workshop on An Application-oriented Approach to BCI out of the laboratory pp 39-42.

[17] Guseva A I, Kireev V S, Bochkarev P V, Kuznetsov I A and Philippov S A 2017 AIP Conference Proceedings 1797 020002

[18] Guseva A I, Kireev V S, Silenko A N, Tikhomirov G V and Sheina E A 2017 AIP Conference Proceedings 1797 020003

[19] Kireev V, Silenko A and Guseva A 2017 IOP Conf. Series: Journal of Physics: Conf. Series 781 012060

[20] Yasnitsky L N, Vauleva S V, Safonova D N and Cherepanov F M 2015 Criminology Journal of Baikal National University of Economics and Law 9 (3) pp 423-430 DOI: 10.17150/1996-7756.2015.9(3) pp 423-430.

[21] Yasnitsky L N and Yasnitsky V 2016 Journal of Applied Economic Sciences XI 8 46 pp 1519-1530

[22] Yasnitsky L N, Dumler A A, Poleschchuk A N, Bogdanov C V and Cherepanov F M 2015 Biology and Medicine (Aligarh) 7 2 BM-095-15 8

[23] Yasnitsky L N, Dumler A A, Bogdanov K V, Poleschuk A N, Cherepanov F M, Makurina T V and Chugaynov S V 2013 Biomedical Engineering 47 3 pp 160-163

[24] Gorchakov A I, Semenova A V, Syrovatskaya Yu V, Shcherbakov Yu V and Yasnitskij L N 2004 Fizika i Khimiya Obrabotki Materialov. 1 pp 43-47

[25] Dobrynin G F and Yasnitskii L N 1995 Glass and Ceramics 51 7-8 pp 275-278

[26] Miloslavskaya N and Tolstoy 2016 Procedia Computer Science 88 pp 300-305

[27] Miloslavskaya N and Tolstoy A 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (Vienna) pp 83-90

[28] Miloslavskaya N and Tolstoy A 2015 Information and Communication Technology 453 pp 135-145