

ИНДЕКСИРАНА БИБЛИОГРАФСКА СПРАВКА ОТНОСНО АНАЛИЗ НА НОРМАЛНИ И АВАРИЙНИ РЕЖИМИ ПО ДАННИ ОТ АВТОМАТИЧНИ РЕГИСТРАТОРИ

Десислав Тодоров

Резюме: В работата е представен процесът на събиране и структуриране на информация за литературни източници, алгоритъм за генериране на индексирана библиографска справка и статистически резултати от анализа на разработената база от данни.

Ключови думи: индексирана библиографска справка, анализ на нормални и аварийни режими

**An indexed bibliography on fault section estimation in electric power plants,
electrical substations, electric transmission and distribution networks**

Desislav Todorov

Abstract: In this work are presented bibliography database preparation, algorithms for indexed bibliography development and statistical results of research.

Keywords: indexed bibliography, fault section estimation

Съвременните електроенергийни обекти са изградени с дълбока степен на автоматизация на производствените процеси. Всички данни от възникнали събития биха могли да бъдат съхранявани и използвани за последващ анализ на процеси и явления. Понякога получената информация е непълна и/или съдържа огромен брой записи, чиято обработка отнема значителен период от време. За решаване на тези задачи са разработени различни прогностични алгоритми, чието систематизиране е цел на настоящото литературно изследване.

Първоначалната библиографска справка, за настоящата разработка, бе изградена по методологията разработена от проф. Николов, стр. 34, [1]. При изпълнение на първи етап бяха посетени посочените в [1] бази от данни, базата от данни на IEEE и бе извършено търсене по ключова дума в приложението Google Scholar. Всички търсения бяха извършени единствено по ключова дума – специфична за представеното изследване. Първоначалната база от данни бе изградена от около 2000 библиографски записа. Ресурсното време, за завършване на този етап, бе около 8 работни часа.

В изследването бе взето решение за използване на софтуер за управление на библиографски записи. След анализ на наличните възможности бе избрана програмата JabRef поради следните особености:

- Софтуерът е безплатен;
- Възможност за работа с BibTex библиографски записи;
- Платформена независимост на продукта – възможност за работа под Windows и Linux;
- Възможност за директно търсене в различни библиографски бази данни;
- Възможност за обработка на библиографска база данни без необходимост от интернет връзка.

Безплатният софтуер Dosear интегрира в себе си няколко програмни пакета, включително JabRef. Съществено негово предимство е визуализация на информацията във вид на мисловни карти. За съжаление работеща версия на този продукт бе публикувана след 2012 година.

JabRef даде много добра възможност за първоначална обработка на базата данни и автоматично откритите повтарящи се записи бяха отстранени. Последващото търсене на дублирани публикации продължи с автоматизирано лексикографско сортиране по автор и статия. Приоритет при ръчното отстраняване на записите бе даден на базите данни SCOPUS и IEEE. При наличие на еднакъв запис, в двете бази данни, той не бе изтриван с оглед на последващата задълбочена обработка на резултатите. След отстраняване на повтарящите се записи бе извършено синтезиране на библиографска база данни, строго насочени към анализа на нормални и аварийни режими в електроенергийни обекти, по заглавия на статии. Описаните в този етап действия отнеха около 40 работни часа. В резултат на извършената работа базата данни бе редуцирана до около 300 записа.

Така разработената база от данни бе подложена на детайлна обработка, включваща:

- Сравнение на дублираните в IEEE и SCOPUS записи по пълнота на информацията;
- Прочитане на резюмета, на публикации, и преценка за целесъобразност по изследваната тематика
- Проверка за достоверност на записи, непубликувани в IEEE база данни и без идентификатор на цифров обект (DOI номер);
- Разработване пълни на библиографски записи съгласно BibTex формат.

Посочената стъпка бе осъществена за около 20 работни дни. Основна част от дейността бе свързана с проверка на достоверността и разработването на пълни записи. Окончателната база от данни бе редуцирана до 219 записа.

Следващ етап в изследването бе получаване на достъп до пълните текстове на публикациите. За целта, първо, бе извършена проверка за наличие на свободен достъп до записите. Разработването на библиографската база от данни, при използване на представената последователност, бе осъществено за около 3 работни месеца.

Поради ограничения финансов ресурс, бе извършено приоритизиране на публикациите и тяхното закупуване. Като резултат базата данни бе запълнена с препратки към 161 файла, с което работата по изграждане на библиографската база от данни бе преустановена.

Всички посочени дотук етапи бяха извършени за около 6 календарни месеца, като в този период е включено самообучението при работа с инсталирания софтуер и различните интернет платформи.

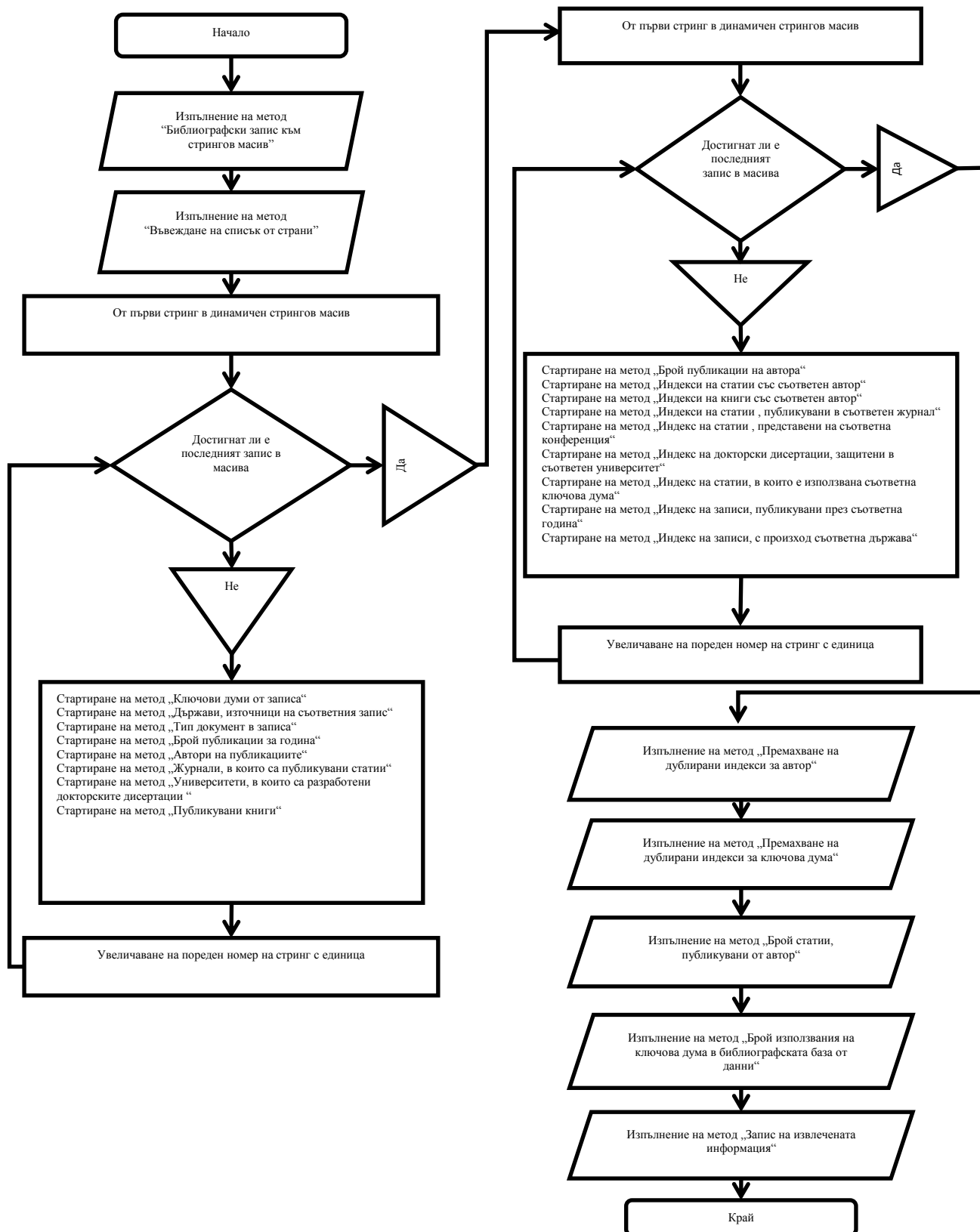
За разработване на индексирания библиографска справка бе използван шаблонът, разработен и приложен от проф. Аландер [2], в повече от 80 публикации. Използваната форма е съставена от два раздела:

- Статистическа информация
 - Тип публикация
 - Разпределение на публикациите по години
 - Класификация по ключови думи
 - Автори
 - Географско разпределение
 - Изводи и заключения
- Индекс
 - Книги
 - Публикации в журналы
 - Дипломни работи
 - Докторски дисертации
 - Магистърски дипломни работи
 - Доклади
 - Патенти
 - Автори
 - Ключова дума
 - Индексиране по години
 - Географски индекс

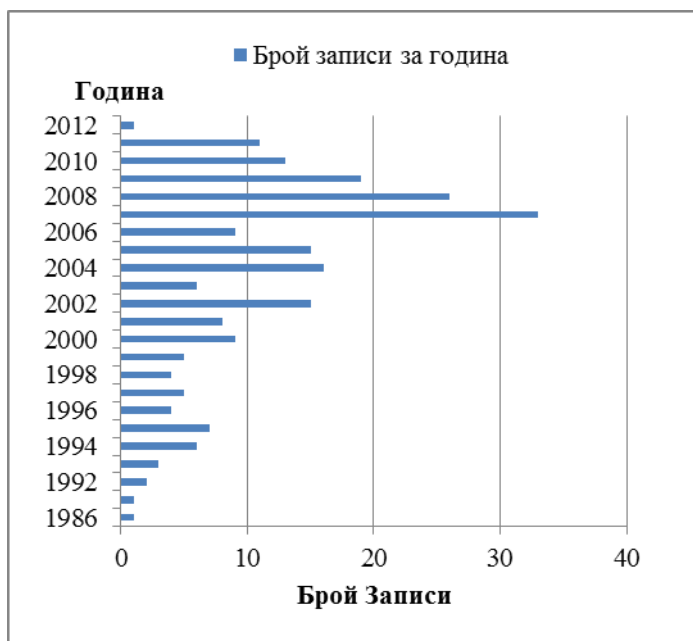
Поради големият обем информация бе разработен софтуерен продукт за генериране на търсените резултати. По същество бе необходимо да се извлече информация от стрингови масиви с полета, представени в Таблица 1

Полета в различни типове библиографски записи			
Статия	Статия на конференция	Докторска дисертация	Книга
<pre>@Article{ bibkey, Title= {}, Author= {}, Journal = {}, Year = {}, Note = {}, Number = {}, Pages = {}, Volume = {}, Abbrev_source_title = {}, Abstract = {}, Affiliation = {}, Author_keywords = {}, Coden = {}, Correspondence_ad dress = {}, Document_type = {}, Doi = {}, ISSN = {}, Keywords= {}, Language = {}, Owner= {}, References = {}, Source = {}, Timestamp = {}, Url = {} }</pre>	<pre>@Conference{ bibkey, Title = {}, Author = {}, Booktitle = {}, Year = {}, Address = {}, Note = {}, Number = {}, Pages = {}, Volume = {}, Abbrev_source_title= {}, Abstract = {}, Affiliation = {}, Author_keywords = {}, Correspondence_address = {}, Document_type = {}, File = {}, Journal = {}, Keywords = {}, Language = {}, Owner = {}, References = {}, Source = {}, Sponsors = {}, Timestamp = {}, Url = {} }</pre>	<pre>@PhdThesis{ bibkey, Title = {}, Author = {}, School = {}, Year = {}, Abstract = {}, File = {}, Owner= {}, Timestamp = {} }</pre>	<pre>@Book{ bibkey, Title= {}, Author= {}, Editor= {}, Publisher = {}, Year= {}, Abstract= {}, File={}, Journal = {}, Owner= {}, Timestamp = {} }</pre>

Таблица 1: Структура на библиографски записи



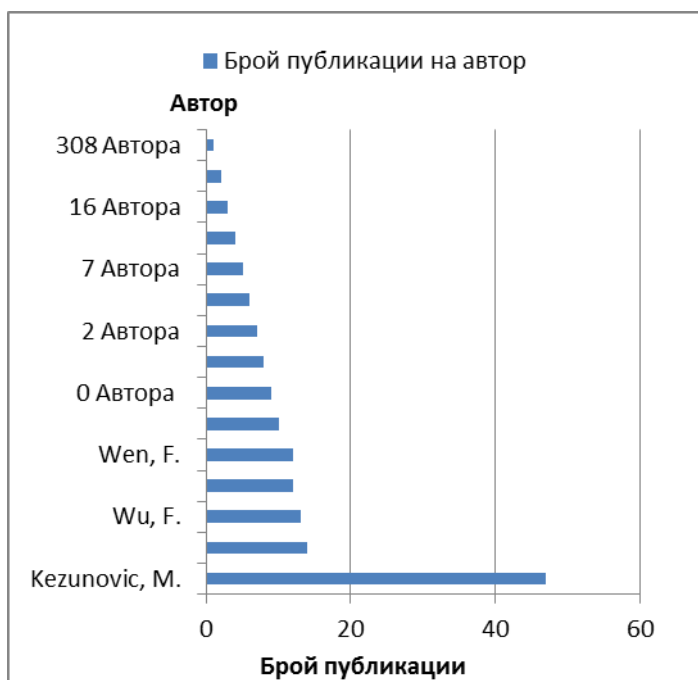
Фигура 1: Алгоритъм на основната програма за генериране на индексирана библиографска справка.



Фигура 2: Публикации по години

Година	Брой записи за година	Година	Брой записи за година
1986	1	2002	15
1989	1	2003	6
1992	2	2004	16
1993	3	2005	15
1994	6	2006	9
1995	7	2007	33
1996	4	2008	26
1997	5	2009	19
1998	4	2010	13
1999	5	2011	11
2000	9	2012	1
2001	8		

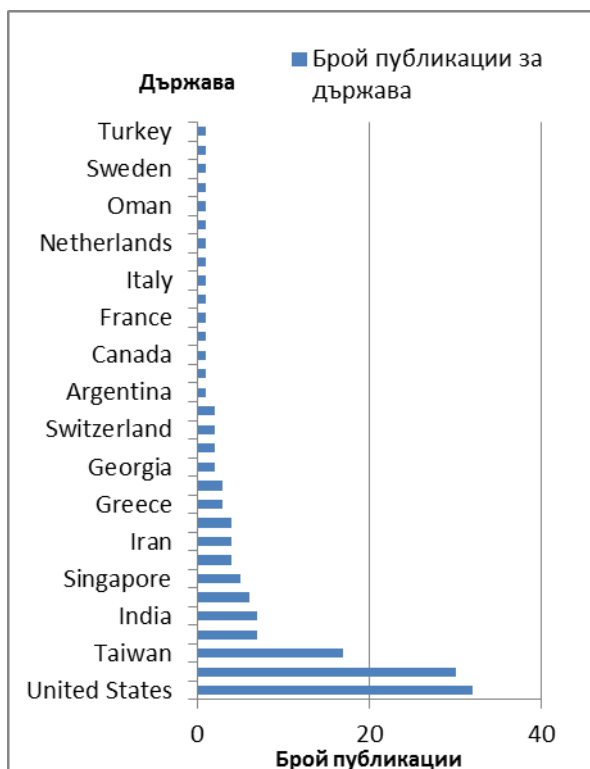
Таблица 2: Публикации по години



Фигура 3: Най - продуктивни автори по тематиката

Автор	Брой публикации
Kezunovic, M.	47
Bi, T.	14
Wu, F.	13
Ni, Y.	12
Wen, F.	12
Wu, F. F.	10
0 Автора	9
2 Автора	8
2 Автора	7
1 Автор	6
7 Автора	5
10 Автора	4
16 Автора	3
58 Автора	2
308 Автора	1

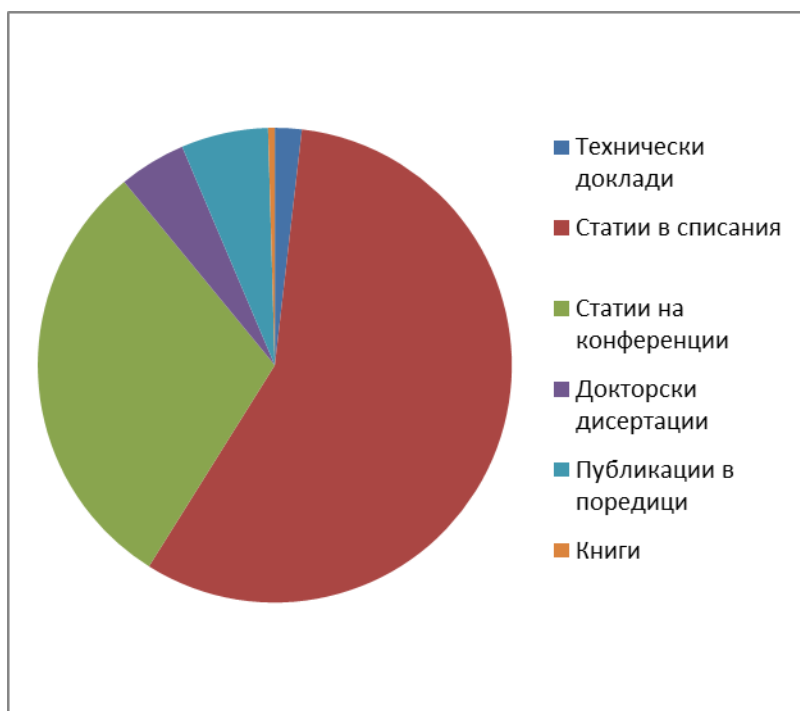
Таблица 3: Най - продуктивни автори по тематиката



Фигура 4: Разпределение на публикациите по държави

Държава	Брой публикации	Държава	Брой публикации
United States	34	Argentina	1
China	30	Bangladesh	1
Taiwan	17	Canada	1
Brazil	7	Czech Republic	1
India	7	France	1
Japan	6	Germany	1
Singapore	5	Italy	1
Egypt	4	Mexico	1
Iran	4	Netherlands	1
United Kingdom	4	Norway	1
Greece	3	Oman	1
UK	3	South Africa	1
Georgia	2	Sweden	1
Spain	2	Thailand	1
Switzerland	2	Turkey	1

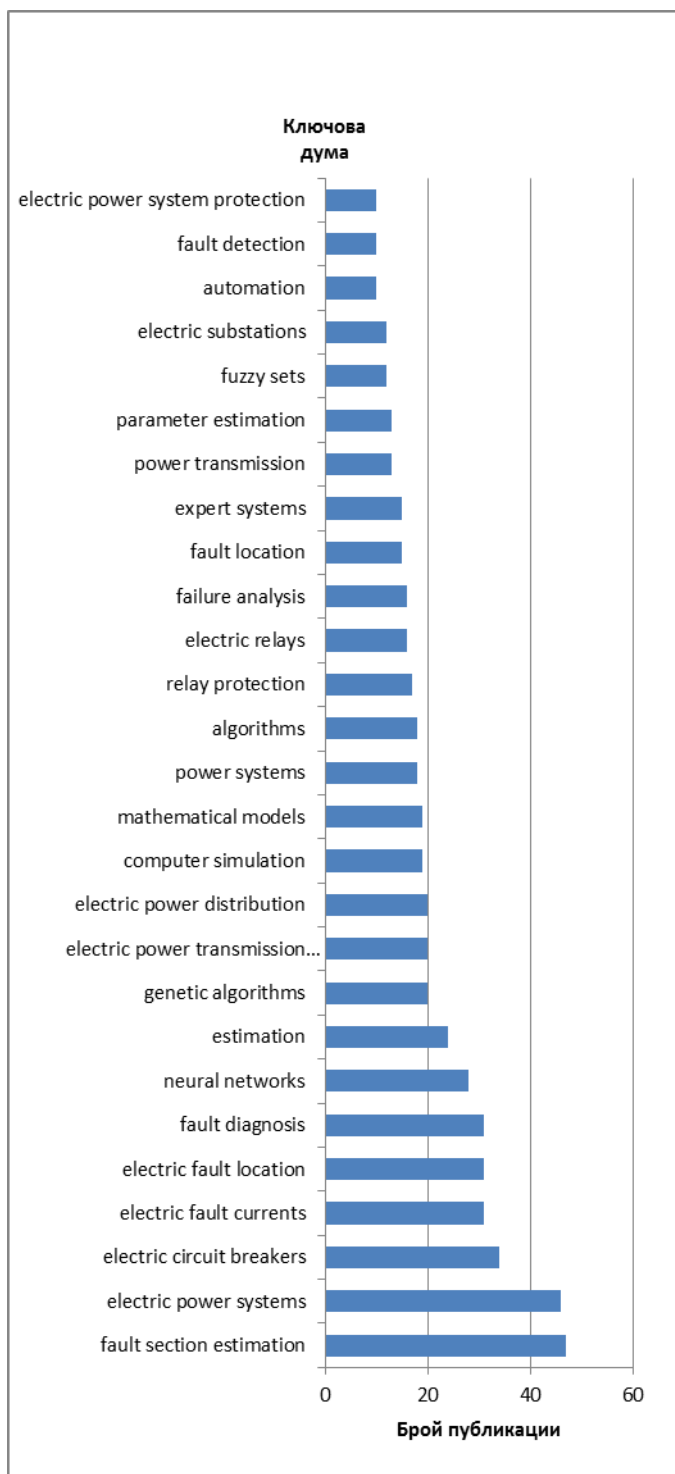
Таблица 4: Разпределение на публикациите по държави



Фигура 5: Дялово разпределение по тип на документи

Тип документ	Брой публикации
Технически доклади	4
Статии в списания	125
Статии на конференции	66
Докторски дисертации	10
Публикации в поредици	13
Книги	1

Таблица 5: Брой публикации за тип документ



Фигура 6: Брой посочвания на ключова дума в публикация

Ключова дума	Брой посочване на ключова дума в публикация	Ключова дума	Брой посочване на ключова дума в публикация
fault section estimation	47	algorithms	18
electric power systems	46	relay protection	17
electric circuit breakers	34	electric relays	16
electric fault currents	31	failure analysis	16
electric fault location	31	fault location	15
fault diagnosis	31	expert systems	15
neural networks	28	power transmission	13
estimation	24	parameter estimation	13
genetic algorithms	20	fuzzy sets	12
electric power transmission networks	20	electric substations	12
electric power distribution	20	automation	10
computer simulation	19	fault detection	10
mathematical models	19	electric power system protection	10
power systems	18	Други ключови думи	1091

Таблица 6: Брой посочвания на ключова дума в публикация

Представеният, на Фигура 1, алгоритъм е разработен на езика C# и програмна среда Microsoft Visual Studio. Програмният продукт е изграден от 537 реда код. Използваните структури от данни са “динамичен масив” и “речник”. Поради необходимостта от лексикографска подредба на резултатите при поредната преработка на софтуера бе използвана структура от данни “подреден речник”. В резултат времето за изпълнението на програмата се увеличи близо пет пъти – до 5 секунди. След запълване на структурите от данни съществена част от работата на алгоритмите бе свързана с търсене на данни в „речник“, по ключ. Установи се, че причина за значителното увеличаване на времето за изпълнение на програмата е частичното използване на структура „балансирано двоично дърво“(“подреден речник”) вместо хеш таблица(“речник”) [3], което ще се вземе предвид при следващата версия на софтуера.

При изграждане на библиографската база от данни и нейната обработка не бе обърнато внимание на различията в стилового изписване на авторите на публикациите. За правилната работа на програмата се наложи ръчно преработване на голяма част от записите.

След първоначален анализ на резултатите се установиха повторения при индексирването на автори и ключови думи, например „fault“ и „fault section estimation“, като този недостатък на програмата в последствие бе отстранен. Времето, необходимо за разработване и тестване на софтуерния продукт бе около 30 календарни дни.

Резултатът от използване на разработената програма е автоматично генерирана библиографска справка в обем около 80 печатни страници. В настоящата публикация е представена получената, от изследването, статистическа информация.

На Фигура 2 и в Таблица 2 са представени публикуваните по тематиката записи, по години. При анализа на резултатите е необходимо да се отбележи, че в библиографската база от данни са включени всички записи, до март 2012 година.

На Фигура 3 и в Таблица 3 са показани авторите, имащи публикации по тематиката. С най – голям брой авторство и съавторство в статии е Младен Кезунович – 47 публикации.

По данните на Фигура 4 и в Таблица 4 може да се предположи, че страните, инвестиращи най - много средства в разработване на сложни експертни системи, прилагани в електроенергийни обекти, са САЩ, Китай и Тайван.

Дяловото разпределение на публикациите е видно на Фигура 5 и Таблица 5. Най – голям е броят на статиите – общо 191. Големият брой публикации и защитените 10 докторски дисертации са показател за значителна работа по разработване на такъв тип експертни системи в световен мащаб.

При анализа на данните от Фигура 6 и Таблица 6 може да се твърди, че основните алгоритми, използвани за анализ на нормални и аварийни режими, по данни от автоматични регистратори са:

- Невронни мрежи – 28 публикации;
- Генетични алгоритми – 20 публикации;
- Размити множества – 12 публикации.

Събирането на библиографска база от данни се оказва продължителна трудоемка задача, изискваща търпение, упоритост и постоянство. Разработването на програмния продукт изискваше относително начални умения по програмиране. Полученият резултат е приемлив, но частта „графично представяне на резултатите“ е разработена с допълнителен готов програмен продукт.

Такъв тип изследвания дават ясна и систематизирана статистическа информация, която би могла да се използва при аргументирано взимане на решения в последващи разработки. Възможно развитие на софтуера би могъл да бъде в посока автоматизирано откриване на първоначалните източници на информация и графично представяне на резултатите.

ЛИТЕРАТУРА

- [1] Е. Николов и Н. Г. Николова, Методология за разработване на научни статии, София: Технически Университет - София, 2013, р. 34.
- [2] J. . T. Alander, „An Indexed Bibliography of Genetic Algorithms in Power Engineering,“ University of Vaasa, Vaasa, 2014.
- [3] С. Наков, В. Колев и колектив, Въведение в програмирането със С#, Велико Търново: Фабер, 2011.

Автори: Десислав Тодоров, главен асистент в катедра „Електроенергетика“, Електротехнически Факултет, ТУ – София, *email:* desislav_todorov@tu-sofia.bg