



УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА У
НОВОМ САДУ



Мирослава Дражић

Прилог решењу аутоматске обраде патентних докумената

МАГИСТАРСКИ РАД

Нови Сад, 2012



УНИВЕРЗИТЕТ У НОВОМ САДУ • ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА
21000 НОВИ САД, Трг Доситеја Обрадовића 6

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Редни број, РБР:	
Идентификациони број, ИБР:	
Тип документације, ТД:	Монографска документација
Тип записа, ТЗ:	Текстуални штампани материјал
Врста рада, ВР:	Магистарски рад
Аутор, АУ:	Мирослава Дражић
Ментор, МН:	проф. др Драган Кукољ
Наслов рада, НР:	Прилог решењу аутоматске обраде патентних докумената
Језик публикације, ЈП:	Српски / латиница
Језик извода, ЈИ:	Српски
Земља публикација, ЗП:	Република Србија
Уже географско подручје, УГП:	Војводина
Година, ГО:	2012
Издавач, ИЗ:	Ауторски репринт
Место и адреса, МА:	Нови Сад; трг Доситеја Обрадовића 6
Физички опис рада, ФО: (поглавља/страна/ цитата/табела/слика/графика/прилога)	<8 поглавља / 84 стране / 30 референци / 15 табела / 52 слике>
Научна област, НО:	Електротехника и рачунарство
Научна дисциплина, НД:	Рачунарска техника
Предметна одредница/Кључне речи, ПО:	Смањење димензионалности, груписање патената, визуелизација
УДК	
Чува се, ЧУ:	У библиотеци Факултета техничких наука, Нови Сад
Важна напомена, ВН:	
Извод, ИЗ:	Ова магистарска теза се бави истраживањем у области аутоматске обраде патентних докумената. Циљ тезе је развој решења које је комерцијално применљиво и представља прилог решењу аутоматске евалуације и класификације патената. У тези се анализирају релевантни аспекти проблематике, предлаже се решење у складу са постављеним циљевима и користе се објективне мере за мерење перформанси појединих делова решења.
Датум прихватања теме, ДП:	24.04.2012
Датум одбране, ДО:	18.12.2012.
Чланови комисије, КО:	Председник: др Миодраг Темеринац, редовни професор Члан: др Мирослав Поповић, редовни професор Члан: др Јован Ђорђевић, редовни професор Члан: др Никола Теслић, редовни професор Члан, ментор: др Драган Кукољ
	Потпис ментора



KEY WORDS DOCUMENTATION

Accession number, ANO :	
Identification number, INO :	
Document type, DT :	Monographic publication
Type of record, TR :	Textual printed material
Contents code, CC :	Master Thesis
Author, AU :	Miroslava Drazic
Mentor, MN :	Dragan Kukulj, PhD
Title, TI :	Contribution to the solution of automatic processing of patent documents
Language of text, LT :	Serbian
Language of abstract, LA :	Serbian
Country of publication, CP :	Republic of Serbia
Locality of publication, LP :	Vojvodina
Publication year, PY :	2012
Publisher, PB :	Author's reprint
Publication place, PP :	Novi Sad, Dositeja Obradovica sq. 6
Physical description, PD : (chapters/pages/ref./tables/pictures/graphs/appendixes)	<8 chapters / 84 pages/ 30 references / 15 tables / 52 pictures >
Scientific field, SF :	Electrical Engineering
Scientific discipline, SD :	Computer Engineering, Engineering of Computer Based Systems
Subject/Key words, S/KW :	Dimensionality reduction, Patent clustering, Visualization
UC	
Holding data, HD :	The Library of Faculty of Technical Sciences, Novi Sad, Serbia
Note, N :	
Abstract, AB :	This master thesis deals with research in the field of automatic processing of patent documents. The aim of the thesis is to develop a solution that is commercially applicable. It represents the contribution to the solution of automated decision support and patent evaluation and classification. The thesis analyzes the relevant aspects of the technology field, proposes a solution in line with the defined goals. Objective measures are used to measure the performance of certain parts of the solution.
Accepted by the Scientific Board on, ASB :	24.04.2012
Defended on, DE :	18.12.2012.
Defended Board, DB :	President: Professor. Miodrag Temerinac, PhD
	Member Professor. Miroslav Popovic, PhD
	Member Professor. Jovan Djordjevic, PhD
	Member: Professor Nikola Teslic, PhD
	Member, Mentor: Professor Dragan Kukulj, PhD
	Mentor's sign

Захвалност

Захвалност дугујем свим драгим људима који су били уз мене и пружали ми подршку онда када ми је подршка била потребна.

САДРЖАЈ

1. Увод.....	1
2. Стање технике	5
2.1 Преглед литературе.....	5
2.1.1 Прикупљање података.....	6
2.1.2 Редукција више димензионалних простора	6
2.1.3 Груписање података	7
2.2 Преглед релевантних патената за предмет истраживања	8
2.3 Преглед техничких решења.....	9
3. Патенти	15
3.1 Судски спорови	16
3.2 Издавање лиценци.....	17
3.3 Вредновање патената	17
3.4 Допринос патената у укупној вредности компаније.....	20
3.5 Патенти као извори информација.....	22
3.6 Структура патента.....	23
4. Развој техничког решења	26
4.1 Прикупљање и складиштења патентних докумената.....	26
4.2 Обрада патентних докумената	27
4.2.1 Анализа патентног текста	27
4.2.2 Редукција података	27
4.2.3 Класификација	28
4.2.3.1 Метода за класификацију на бази кодова	31
4.2.3.2 Метода за класификацију на бази кодова и технологија.....	33

4.3	Визуализација	34
4.4	Скуп Података	35
5.	Програмска подршка - ПСАЈМ	36
6.	Резултати и анализа	44
6.1	Перформансе интернет робота.....	44
6.2	Верификација расподеле патената у 2Д простору	45
6.3	Тест метода за класификацију података	46
6.3.1	Опште познате методе.....	46
6.3.1.1	Визуелно поређење	47
6.3.1.2	Математичко поређење.....	48
6.3.2	Властите методе	49
6.3.2.1	Тест методе „груписање по технологијама“	49
6.3.2.1.1	Математичко поређење	50
6.3.2.1.2	Визуелно поређење	51
6.3.2.2	Тест методе „груписање по кодовима“	54
6.3.2.2.1	Математичко поређење	54
6.3.2.2.2	Визуелно поређење	55
6.3.2.3	Закључак.....	58
6.4	Употребна вредност алата у реалним случајевима.....	60
6.4.1	Оцена јачине портфолија компаније.....	60
6.4.2	Анализа судских спорова.....	63
6.4.2.1	Патенти који су укључени у спорове против Андроид-а	66
6.4.2.2	Класификација патената по технологији на коју се односе	68
6.4.2.3	Класификација патената по власнику патентног права.....	69
6.4.2.4	Класификација патената методом „груписање по кодовима“	72
6.4.2.5	Анализа кључних патената.....	73
6.4.2.5.1	Патент US5,946,647.....	74
6.4.2.5.2	Патент US6,343,263.....	75
6.4.2.5.3	Патент US6,125,447.....	76
6.4.2.6	Патенти у власништву компаније “Motorola”	77
6.4.2.7	Закључак.....	79
7.	Закључак	80
8.	Литература.....	82

СПИСАК СЛИКА

Слика 1: Изглед прозора за претрагу алата „ <i>Matheo Patent</i> ”	10
Слика 2: Главни прозор апликације „ <i>Patent iNSIGHT Pro</i> “	11
Слика 3: Један пример визуализације резултата за „ <i>Patent iNSIGHT Pro</i> “	12
Слика 4: Пример визуализације „ <i>Data Validation</i> ” алата.....	13
Слика 5: Изглед „ <i>CobaltIP</i> “ апликације са отворена два прозора.....	13
Слика 6: Визуализација резултата „ <i>CobaltIP</i> “ апликације.....	14
Слика 7: Састав корпоративних добара	21
Слика 8 Изглед прве странице патентног документа.....	24
Слика 9 Алгоритам методе за класификацију.....	32
Слика 10 Основни модел програмске подршке	36
Слика 11 Генерална поставка система	37
Слика 12 Пример изгледа главног прозора	38
Слика 13 Случај употребе	39
Слика 14 Изглед прозора програмске подршке	41
Слика 15 Изглед додатних информација о патенту.....	43
Слика 16 Верификација расподеле патената у 2Д простору	46
Слика 17 Резултат експерата и <i>Ronn</i> алгоритма	47
Слика 18 стручна анализа наспрам <i>k-means</i> алгоритма	48
Слика 19 Резултат примене методе „груписање по технологијама“	49
Слика 20 Приказ односа тачака у групи и тачака издвојених из групе	51
Слика 21 Позиционирање групе „ <i>Cryptography/Security</i> “	52
Слика 22 Позиционирање групе „ <i>Software Data Management</i> ”	52
Слика 23 Позиционирање групе „ <i>Smart Homes</i> ”	53

Слика 24	Позиционирање групе „Wireless”	53
Слика 25	Резултат примене методе „груписање по кодовима“	54
Слика 26	Приказ односа тачака у групи и тачака издвојених из групе	55
Слика 27	„G02B“ – уређаји, системи и елементи везани за оптику	56
Слика 28	„G06F“ - дигитални процесори за обраду података	57
Слика 29	„H04L“ – пренос дигиталних информација	57
Слика 30	„H04N“ - телевизија	58
Слика 31	Резултат груписања на основу првог дела кода	59
Слика 32	Груписање по технологијама	60
Слика 33	19 патената груписаних по кодовима до трећег нивоа	61
Слика 34	19 патената у простору са 19 патената компаније „Microsoft“	62
Слика 35	19 патената у простору са 726 патената компаније „Microsoft“	62
Слика 36	Спорови и лиценце у вези са Андроид оперативним системом	64
Слика 37	Дистрибуција патентних спорова	68
Слика 38	Дистрибуција патентних спорова приказана ПСАЈМ алатом	68
Слика 39	Расподела патената који се односе на Андроид ОС	69
Слика 40	Расподела патената који се односе на Андроид ОС по технологији	69
Слика 41	Расподела патената у спору приказана по власницима патентних права	70
Слика 42	Расподела патената по власнику патентног права - Microsoft	71
Слика 43	Расподела патената по власнику патентног права – BT	71
Слика 44	Литигациони патенти и патенти компаније „Google“	72
Слика 45	Груписање по кодовима	73
Слика 46	Патент US5,946,647	74
Слика 47	Патент US6,343,263	75
Слика 48	Патент US6,125,447	76
Слика 49	Три анализирана патента у 2Д простору	77
Слика 50	Сви цитати патената у спору	78
Слика 51	Патенти који цитирају (лево) и патенти који су цитирани (десно)	78
Слика 52	Патенти компаније „Motorola“ у односу на литигационе	79

СПИСАК ТАБЕЛА

Табела 1 Типови података у појединим деловима патентног документа.....	25
Табела 2 Пример сортирања патената по кодовима	32
Табела 3 Пример пребројавања	33
Табела 4 Резултат алгоритма	33
Табела 5 Структура табеле за улазне податке	41
Табела 6 Библиографски подаци	42
Табела 7 Статистика прикупљених података за први тестни случај	44
Табела 8 Брзина прикупљања једног патената за први тестни случај	45
Табела 9 Други тест случај 1820 патената	45
Табела 10 Тест случај 1820 патената просечно	45
Табела 11 Тачност класификације четири алгоритама груписања	48
Табела 12 Перформансе алгоритама груписања	48
Табела 13 Провера тачности кластер методе	50
Табела 14 Провера тачности кластер методе	55
Табела 15 Приказ парница усмерених против Андроид ОС-а	67

СКРАЋЕНИЦЕ

IPR	- <i>Intellectual Property Rights</i>
IPC	- <i>International Patent Classification</i>
PID	- <i>Patent Identification</i>
GDP	- <i>Gross Domestic Product</i>
SME	- <i>Small-Medium Enterprise</i>
HTML	- <i>Hypertext Markup Language</i>
CSV	- <i>Comma Separated Values</i>
OS	- <i>Operating System</i>
US	- <i>United States</i>
EPO	- <i>Europien Patent Office</i>
USPTO	- <i>United States Patent and Trademark Office</i>
TF-IDF	- <i>Text Frequency – Inverse Document Frequency</i>
MDS	- <i>Multidimensional Scaling</i>
RONN	- <i>Re-Organizing Neural Network</i>
FCM	- <i>Fuzzy c-means Clustering</i>

1. Увод

Интернет као највећи и најсвеобухватнији извор информација постаје саставни део живота савременог друштва. Међутим, управо та огромна количина доступних информација намеће потребу за претраживањем и издвајањем само оних информација које су од интереса за корисника интернет мреже и сортирањем тих информација по жељеним критеријумима. Проблем претраживања интернета, чак и са многим постојећим претраживачима, попут "*google.com*" далеко је од прихватљивог, поготово за претраживања података који захтевају анализу великог броја докумената (речи, информација) по специфичним условима. Због великог броја различитих потреба претраживања интернета, корисници су принуђени да развијају властита, или надограде већ постојећа решења како би осигурали што бржи и тачнији приступ искључиво жељеним информацијама. Са друге стране, и када се издвоје само жељене информације из огромног скупа доступних информација, и даље тај скуп може бити довољно велик да га је немогуће довољно добро анализирати и обрадити у релативно кратком временском року. Из наведених разлога намеће се потреба за алатима чија би улога била аутоматска анализа, обрада и презентовање података корисницима. Аутоматска анализа текстова патентних портфолија, састављених често од великог броја патената, има за циљ стицање суштинског разумевања технолошких карактеристика компаније, као и стања и трендова у разматраној индустрији и тржишту.

У доба све веће конкурентности на тржишту и примата компанија које врло брзо усвајају нова техничка решења јавља се све већа потреба за бржим технолошким трансфером и у ту сврху и за евалуацијом постојећих техничких решења. Овај тренд је поготово важан за развој малих и средњих предузећа која кроз патентирање, а потенцијално и лиценцирањем својих технолошких решења могу равноправно да конкуришу великим мулти-националним компанијама. У ту сврху поступак анализе

технологија има за циљ да помогне у суштинском разумевању стања и трендова, тржишта и конкуренције у разматраној технолошкој области. Да би анализа технологија и трендова била могућа неопходно је обезбедити приступ великој количини разнородних података који су јавно доступни, у данашње време најчешће путем интернета. Веома значајан део у анализи технологија се односи на анализу интелектуалне својине, као што су патенти. Последњих година ова област је у великој експанзији и постоји изузетно интересовање и потреба за новим алатима који би помогли у решавању проблема вредновања и класификације интелектуалне својине, а посебно патената. Овај рад је резултат потребе да се примене неке од познатих метода у решавању проблема аутоматске обраде патентних докумената, али и да се развију нове методе комбинујући специјализоване алате и људску експертизу у сврху анализе свих патената који су у власништву компаније (тзв. патентни портфолио). Анализа сопствених патентних портфолија омогућава компанији да развије стратегију у смислу развоја нових производа или лиценцирања постојећих решења, док анализа патената конкурентних фирми доприноси бољој процени конкуренције и даје индикацију о смеру њиховог развоја. Неопходност једног оваквог, или сличног, решења биће илустровано реалним и актуелним проблемом пред којим се наша једна од највећих светских компанија – „*Google*“ и тренутним „ратом“ који се води на подручју патената, а усмерен против једног производа ове компаније – Андроид система. Овај, такозвани „рат“ биће детаљно описан у поглављу 6 и на њему ће бити демонстриране неке од могућности техничког решења и изведени закључци на основу аутоматске обраде патената и патентних портфолија.

Циљ овог истраживања је одабир метода тј. алгоритама за анализу текста патената како би се на аутоматизован начин патенти обрадили и класификовали по сличности. Обрада патената између осталог подразумева сажимање високо-димензионалних података, а класификација употребу једне од опште познатих метода или развој сопственог алгорита за одређивање кластера. Документи који су предмет обраде засновани су на оригиналним садржајима патената заведени и јавно доступни на интернету. Две главне базе патената у свету су „*EPO*“ (енгл *European Patent Office*) и „*USPTO*“ (енгл *United States Patent and Trademark Office*).

Патентни документи, као правно-технички документи, садрже прецизне описе метода, система и/или техничких уређаја који територијално штите проналазак и као такви могу бити коришћени само под правно регулисаним условима. Неопходно је направити програмску подршку која може да, као рачунарски алат, пружи подршку у анализи технологија неопходних приликом доношења пословних одлука као што су:

- (i) патентирање техничких решења,

- (ii) нуђење својих патентних решења за лиценцирање другим компанијама
- (iii) куповина/продаја компанија и/или патентних пакета (енгл. *IP pools*), или
- (iv) приликом разрешавања судских спорова узрокованим нелегалним коришћењем интелектуалне својине (патената).

Проблем се своди на прикупљање патентних докумената и аутоматску обраду текста како би се ефикасно вредновао портфолио неке компаније. Како се број патената повећава и количина података расте, немогуће је успешно анализирати сваки скуп патената без груписања. Дакле, груписање је основна функција коју свака алатка за анализу патената треба да обезбеди. Да би се то успешно урадило неопходно је одабрати одговарајући алгоритам за класификацију. У ту сврху треба одабрати неколико метода и тестирати их, али и одредити мере на основу којих ће се проценити која је од тих метода боља, и евентуално развити нови алгоритам.

Очекивани резултати истраживања се могу поделити на практичне и научне резултате. Научни резултати ће бити у форми реализованих алгоритама за редукцију података и класификацију патената, а практични у виду програмске подршке за њихову имплементацију у конкретним проблемима, као сервис (подршка) у анализи интелектуалне својине. За верификацију ће се користити реални примери патентних портфолија одређених компанија. Неки од очекиваних резултата истраживања су:

- интернет оријентисана програмска подршка за прикупљање и пред-обработку патентних докумената - интернет оријентисана апликација за претрагу *USPTO* сајта (као најважнија база података у области америчких патената), повлачење и пред-процесирање патентних докумената (провера редуваности, статистичка анализа, парсирање и структурирање текстуалних и нумеричких података),
- програмска подршка за редукцију података садржаних у тексту патентних докумената заснована на лексичком моделу,
- програмска подршка за анализу редуваног простора на кластере података и визуализацију група тачака по изабраном критеријуму у две димензије (2Д): приказ технологија, патената, компанија, релација цитираности итд.
- генерисање извештаја.

Програмску подршку треба урадити у неколико фаза реализовањем одговарајућих модула:

- A. прикупљање и складиштење података;
- B. обрада података: редукција добијених високо-димензионих података у простор знатно нижег реда (2Д), при чему се тежи да структура података остане што сличнија оригиналној; класификација података након редукције

С. визуализација података и презентација са осталим информацијама

Овај рад даје прилог решењу аутоматске обраде патентних докумената у смислу класификације патената по више критеријума у циљу боље анализе патентних портфолија компанија. У поглављу 2 дат је преглед стања технике из области истраживања у смислу прегледа литературе, патената и примењених техничких решења (алата), у поглављу 3 дат је опис патената, судских спорова, начина вредновања патената и уопште не материјалних добара у предузећу, преглед информација које се могу наћи у патентним документима и структура патентних докумената. У поглављу 4 изложена је методологија истраживачког рада, а поглавље 5 описује развијену програмску подршку названу „ПСАЛМ“ која представља резултат истраживања у облику алата намењеног аутоматској обради патентних докумената. Поглавље 6 представља сублимацију свих резултата добијених објективним и субјективним оценама појединих делова програмске подршке као и оцену њене употребне вредности на реалним патентним портфолијима. У поглављу 7 дат је закључак са освртом на могућност даљег усавршавања и могућим правцима даљег развоја, а у поглављу 8 преглед коришћене литературе.

2. Стање технике

У овом поглављу дат је преглед литературе која је релевантна за област истраживања, преглед патената из дата области који би евентуално могли бити повређени и преглед постојећих алата (примењених техничких решења).

2.1 Преглед литературе

Све више је истраживања која се баве аутоматском обрадом патентних докумената са циљем да кроз анализу и обраду патената извуку што више квалитативних информација. Већина истраживања се бави груписањем патентних докумената на основу неких обележја која су добијена лексичком обрадом, а нека укључују и семантику. Нека иду толико далеко да са аутоматском обрадом патентних докумената покушавају открити чак и одређену вероватноћу да је дати патент нарушен. У истраживању [1] описан је метод који на бази технолошке сличности одређује вероватноћу нарушавања патентног права уз помоћ семантичке обраде текста и тзв. методологије субјекат-акција-објекат (САО).

Овакве и сличне методе укључују неколико неизоставних корака као што су: прикупљање података, редукција више димензионалних простора, груписање података и визуализацију. Прикупљање је неопходан корак пре било какве обраде података јер се подаци од интереса углавном налазе у различитим базама података чак и на различитим серверима. Стога је неопходно прикупити их и ускладиштити у локалну базу података како би били лако доступни. Након прикупљања података такође неопходан корак је и редукција више димензионалног простора у простор ниже димензије како би подаци могли бити лакше презентовани. Простор нижег реда обично је 2Д или 3Д простор. Затим је обично потребно дате податке груписати (класификовати) по неком од критеријума у зависности од тога које се карактеристике скупова података желе истаћи (приказати).

2.1.1 Прикупљање података

Како би све информације о патентима биле доступне за лакшу анализу потребно их је прикупити и ускладиштити. За то су погодни интернет роботи, програми који аутоматски прикупљају информације са интернет страница по унапред дефинисаном правилу, чувајући их ради лакше анализе и тумачења. Ови програми су још познати под називом интернет пауци (енгл *web spiders, web crawlers*).

Постојећа решења интернет робота се могу поделити у 2 врсте – она која имају могућност издвајања корисних информација са интернет странице и она која раде по принципу преузимања читавих презентација. Прва група интернет робота се није показала довољно прилагодљивом страницама које имају лошу структуру *HTML* кода, а друга група је неупотребљива јер не постоји неопходна контрола рада, као и издвајање корисних информација са интернет странице. Из наведених разлога у оквиру овог истраживачког рада развијена је властита програмска подршка која обавља посао интернет робота и која ће бити детаљније описана у поглављу 5.

2.1.2 Редукција више димензионалних простора

Након што се прикупе и ускладиште патентни подаци потребно их је обрадити. Да би се ти мулти-димензионални подаци могли обрадити неопходно је применити технику редукције димензионалности података. Редукција димензионалности података је од велике важности јер решава низ нежељених својстава високо-димензионих простора и омогућава решавање проблема у техничким и пословним системима. Захваљујући развијеним методама редукције високо-димензионих модела могуће је решити захтеве за визуализацијом, класификацијом, моделовањем и управљањем сложеним системима [2]. Постоје два главна приступа у редукцији димензија система: методе које на основу редувантности и релевантности селекују и издвајају доминантна обележја, а остала занемарују (енгл *feature forward selection*) и методе које селекцију обележја врше трансформацијом полазног простора у простор много нижег реда, уз чување основних карактеристика (енгл *PCA, MDS, GHA*). Поступак редукције података је кључан за решавање многих проблема у пословним или техничким системима који прикупљају, преносе или складиште велике количине података.

У раду [3], предлаже се метрика за оцењивање квалитета визуелизације резултата редукције димензионалности простора на основу корелације међусобних удаљености у многострукости података. Предложена метрика је упоређена са другим метрикама и добијени су резултати примене различитих техника за оцену смањења димензионалности

који показују да је предложена метрика погодна за квантитативно вредновање резултата технике смањења димензионалности.

Посебну групу метода за издвајање обележја чине неке врсте вештачких неуронских мрежа (ВНМ) чији је циљ екстракција кључних карактеристика система. Ове методе су иницијално највише развијене и примењиване у области препознавања облика. Оне се заснивају на процесу трансформисања улазног у излазни простор нижег реда. Овде је од посебног интереса *Hebb*-ов закон само-обучавања и *Kohonen*-ов концепт само-организованог пресликавања обележја, односно концепт компететивног учења [4],[5]. Оба концепта представљају технике обуке без надзора.

2.1.3 Груписање података

Кластеризација (или груписање) је подела скупа $\{o = O_1, O_2, \dots, O_n\}$ објеката у хомогене кластере која максимизира сличност унутар кластера, док минимизира сличност између кластера. Кластери су формирани без икакве претходне информације о објектима који су груписани. Све ознаке повезане са објектима се добијају искључиво из података. Груписање помаже идентификовању важних тема или концепата из низа докумената и помаже у препознавању неприметних или неочекиваних образаца уграђених у базама података. То је један од најпопуларнијих и најчешће коришћених приступа за формирање смислене групе из необележених објеката. Груписање је основна техника у многим дисциплинама, укључујући машинско учења и вештачку интелигенцију. Она се нашироко користи у разним процесима доношења одлука.

Циљ кластер анализе је да формира такве групе објеката да објекти у једној групи буду сличнији међусобно него са објектима из других група. Сличност је обично представљена као нека врста дистанце. Најпознатије су тзв. Еуклид-ове дистанце.

У последње време све је чешћа примена неуронских мрежа (математички модели који кроз обучавање могу да препознају структуре у подацима). Неуронске мреже се из области вештачке интелигенције све више селе у област „копања по подацима“ (енгл *data mining*). У кластер анализи су најпознатије Кохененове неуронске мреже, које имају улазни и излазни слој и сви неурони улазног слоја су повезани са свим неуронима излазног слоја [4][5].

У раду [6], предложено је окружење које интегрише два нивоа знања, једно у облику међусобних ограничења и друго у облику атрибута – кључних речи, у методи за класификацију докумената. Прво се тежине поставе на почетне вредности на основу метрике учења са узајамним ограничењима, затим се „уче“ две врсте знања комбиновањем

приступа заснованог на дистанци и приступа заснованог на ограничењима, на крају се процењује резултат изабраних груписања на основу степена задовољства корисника.

Нови приступ у идентификацији основне технологије из перспективе технолошких крос-утицаја на основу информација из патентне класификације узимајући у обзир укупне међусобне односе међу технологијама предложен је у студији [7]. Прво се израчунају технолошки индекси крос-утицаја, а затим се израчунају приоритети технологија уз уважавање њихових директних и индиректних утицаја.

2.2 Преглед релевантних патената за предмет истраживања

Ово поглавље даје кратак преглед патената који су релевантни за дату област истраживања аутоматске обраде и класификације докумената са аспекта прегледа владајућих ставова, али такође и могућности да се властитим решењем не наруши неки од постојећих, важећих патената.

Проналазак „US6654743“ се односи на поље електронског прикупљања докумената, а посебно на формирање кластера електронских докумената на основу неке мере сличности између електронских докумената [8]. Ово одобрено патентно решење дефинише метод и уређај за груписање докумената. Поступак обухвата издавање захтева од електронског уређаја за одређене документе који су релевантни за захтев. Документи за које се сматра да испуњавају услове захтева се идентификују и потом испоручују систему за груписање докумената. Систем груписања докумената затим обрађује дате документе како би формирао групе докумената, тј. кластере. Само-подешавајући процес се затим покреће како би се добили робусни кластери који се затим приказују кориснику. Добијање робусних кластера је омогућено понављањем процеса груписања и употребом варијација на једно од следећег:

- достављање докумената у различитом редоследу,
- примена различитих алгоритама груписања,
- коришћење различитих матрица сличности.

Робусни кластери представљају оне кластере који се најчешће формирају.

Наведени патент представља генерално решење за груписање електронских докумената па би се самим тим могао односити и на патентне документе. Међутим, његова специфичност је у томе што дефинише механизам издвајања робусних кластера поступком понављања процеса груписања што га издваја од осталих техничких решења.

Пријава проналаска „US2009/0070366A1“ се односи на поље груписања интернет (енгл *web*) докумената [9]. Описани су метод и систем који се састоје од следећих корака:

- прикупљања информација о линк-овима и структури директоријума улазних докумената
- извлачење хијерархијске структуре из прикупљених информација
- генерисање и приказивање једног или више кластера интернет докумената на основу хијерархијске структуре

Наведени патент представља решење за груписање интернет докумената, па би у ужем смислу се могао односити и на груписање патената. Његова специфичност је у томе што приликом процеса груписања издваја и користи хијерархијску структуру у информацијама.

2.3 Преглед техничких решења

Разни софтверски алати развијени су у области анализе патената. Они могу анализирати поједине патенте, али и цела патентна портфолија; прикупити патентне информације и направити основне статистичке анализе као и визуелно приказати дате податке. Већина алата користи статистичке методе за анализу података о патентима у одређеном временском раздобљу и приказују трендове у кретањима патената помоћу графова и табела. То су углавном комерцијалне програмске подршке развијене у циљу пружања услуга аутоматске обраде и класификације патената. Међутим, већина патентних база података и доступних алата су скупи и стога нису приступачни за мала и средња предузећа или академске институције, посебно у земљама у развоју. Осим тога, ови алати могу бити компликовани за коришћење или захтевати високу стручност у подручју интелектуалног власништва. Један од комерцијално доступних алата који се бави тражењем, преузимањем и анализом података о патентима је: “*Matheo Patent*” апликација. Између осталог бави се иновацијама, индустријском својином, праћењем патената, истраживањем и развојем, ефикасном експлоатацијом патената, итд. Такође омогућава брз приступ информацијама са масовних патентних база:

- „*Esp@cenet*“, глобална базе патената Европског завода за патенте (енгл *European Patent Office*). Она дозвољава приступ "на више од 70 милиона патентних докумената широм света, који садржи информације о проналасцима и техничким достигнућима од 1836 до данас".
- „*USPTO*“, скраћено од "*United States Patent and Trademark Office*" Канцеларија за патенте и жигове Сједињених Држава, је виђена као најважнија канцеларија у области патената.

„*Matheo Patent*” дозвољава претрагу патента по кључним речима: наслову, резимеу, проналазачима, подносиоцу пријаве патента, патентним бројевима, класификационом

коду (Слика 1). Све тражене информације се аутоматски преузимају у локалну базу података на рачунару. Захваљујући опцији за преузимање, може се преузети, ако је потребно, опис, захтеви, референце цитата, насловна страна и комплетна породица патента [10].

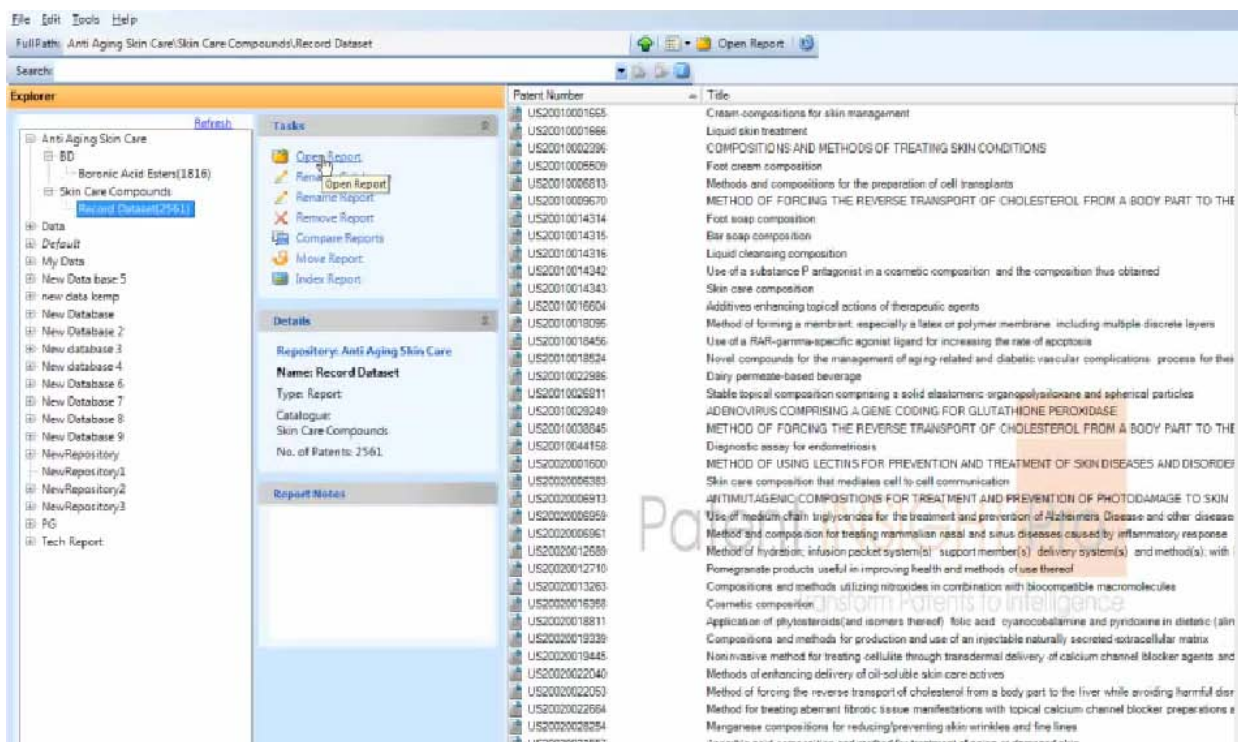
The screenshot shows a web application window titled 'Wizard - New Project - [snowboard_binding]'. The main heading is 'Easy Patent Search'. On the left, there is a search filter sidebar with fields for 'Title or Abstract or Applicant or Inventor', 'Title', 'Inventor', 'Applicant', 'Patent Number', 'Application Number', 'Priority Number', 'IP Classification', 'E Classification', and 'Publication Date (Year)'. The search criteria entered are 'snowboard and binding'. Below these fields are 'Search' and 'Reset' buttons, and a link for 'Advanced Search'. The main search results area shows '1401 results found in the Espacenet Worldwide database for: (TA=snowboard and TA=binding) using Guided Search.' It includes a 'Download Cart (0)' and a 'Download All (1401)' link. The results are displayed in a table with columns for 'Patent' and 'Title'. The first few results are:

Patent	Title
<input type="checkbox"/> EP2414057A1	Snowboard binding having rear entry and asymmetrical leg support 2012-02-08 SAM SPORT AND MARKETING AG [CH]
<input type="checkbox"/> WO2012009282A2	Strap for snowboard binding 2012-01-19 K 2 CORP [US]
<input type="checkbox"/> US2012007339A1	Strap for snowboard binding 2012-01-12 K 2 CORP [US]
<input type="checkbox"/> US2011316256A1	Modular binding for sports board 2011-12-29
<input type="checkbox"/> GB2481685A	Suspension apparatus for a skateboard truck 2012-01-04 STREETBOARDZ HOLDINGS PTY LTD [AU]
<input type="checkbox"/> WO2011158977A1	Snowboard binding strap 2011-12-22 BUZRUN CO LTD [KR]
<input type="checkbox"/> WO2011155845A1	Snowboard 2011-12-15 HITURN AS [NO]
<input type="checkbox"/> EP2385865A2	Snowboard binding with lower impact on board flex 2011-11-16 BURTON CORP [US]
<input type="checkbox"/> EP2380639A1	Snowboard binding 2011-10-26 BUZRUN CO LTD [KR]
<input type="checkbox"/> US2011254251A1	Snowboard binding 2011-10-20 BUZRUN CO LTD [KR]
<input type="checkbox"/> GB2478528A	Rotatable binding mount for snowboard 2011-09-14 COSGROVE JAMES [GB]
<input type="checkbox"/> CA2748210A1	Snowboard binding 2010-07-01 BUZRUN CO LTD [KR]
<input type="checkbox"/> US2011204597A1	Safety binding for a snowboard 2011-08-25 UNIV MAGDEBURG TECH [DE]

At the bottom of the window, there are 'Back to Search Method' and 'Close' buttons.

Слика 1: Изглед прозора за претрагу алата „*Matheo Patent*”

Још један од комерцијалних алата из ове области је тзв. „*Patent iNSIGHT Pro*“ (Слика 2). То је свеобухватна платформа за истраживање и анализу патената која значајно убрзава време потребно за анализу патената. Укључује сложене алгоритме за анализу великог броја текстуалних података (енгл *Text Mining*) како би извео закључке у једној минути који би иначе за истраживача трајали данима. Дакле, помоћу овога алата, може се спровести истраживање у минутима и сатима уместо у данима и месецима.



Слика 2: Главни прозор апликације „Patent iNSIGHT Pro“

Алат може бити користан при доношењу одлука и одређивању стратегија, а све на основу добрих и поузданих информација. Може да генерише професионалне и високо-стручне извештаје који утичу на креирање стратегије предузећа. Поједностављује комуникацију између тимова и омогућава извештавање у оквиру организације по стандардима индустријске праксе у области патената. Такође умањује шансе за људске грешке које се могу десити у мануелним процедурама [11].

Оно што је специфично за овај алат је то што поред информација о патентима прикупља и податке из журнала, тј. часописа. То му омогућава одређену ширину приликом интерпретације резултата. Слика 3 показује један пример визуализације резултата.



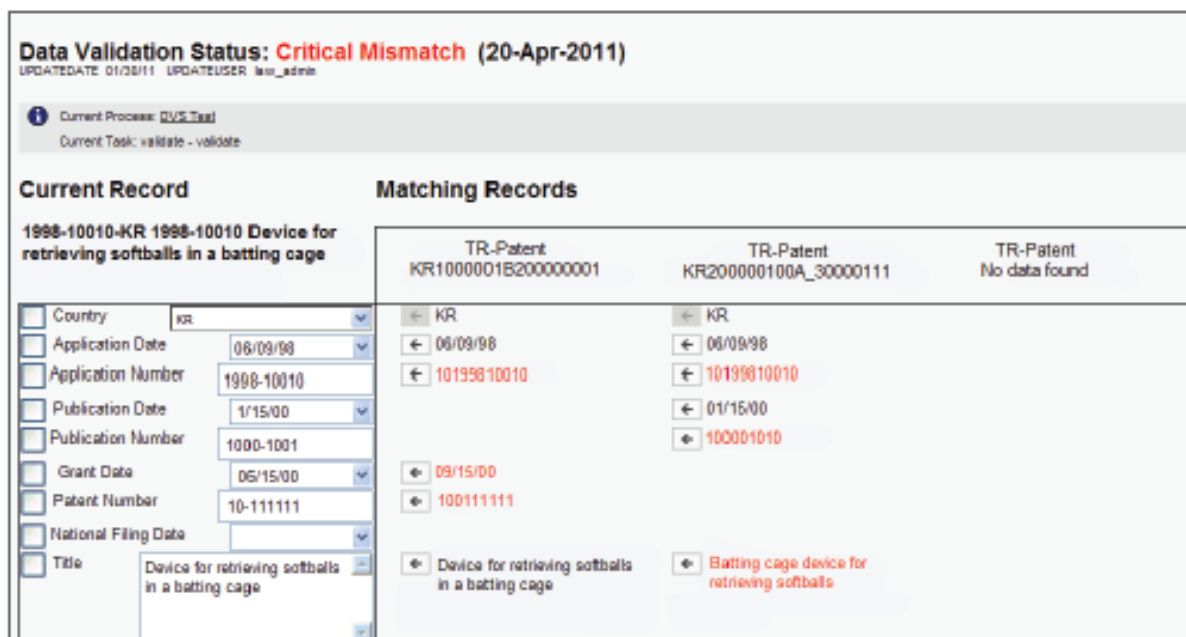
Слика 3: Један пример визуализације резултата за „Patent iNSIGHT Pro“

„Thomson IP Management Services“ су ресурси који пружају могућност да се из једног патентног портфолија извуче максимална вредност. „Thomson IP Manager“ [12] је алат за управљање средствима интелектуалне својине на нивоу предузећа које је део пакета „Thomson IP Management Services“. Овај алат пружа могућност различитим одељењима у предузећу да међусобно сарађују, унапређује процес доношења одлука и слично. „Thomson IP Manager“ нуди још два изузетно корисна алата у свом склопу. То су:

- *Process Architect*
- *Data Validation*

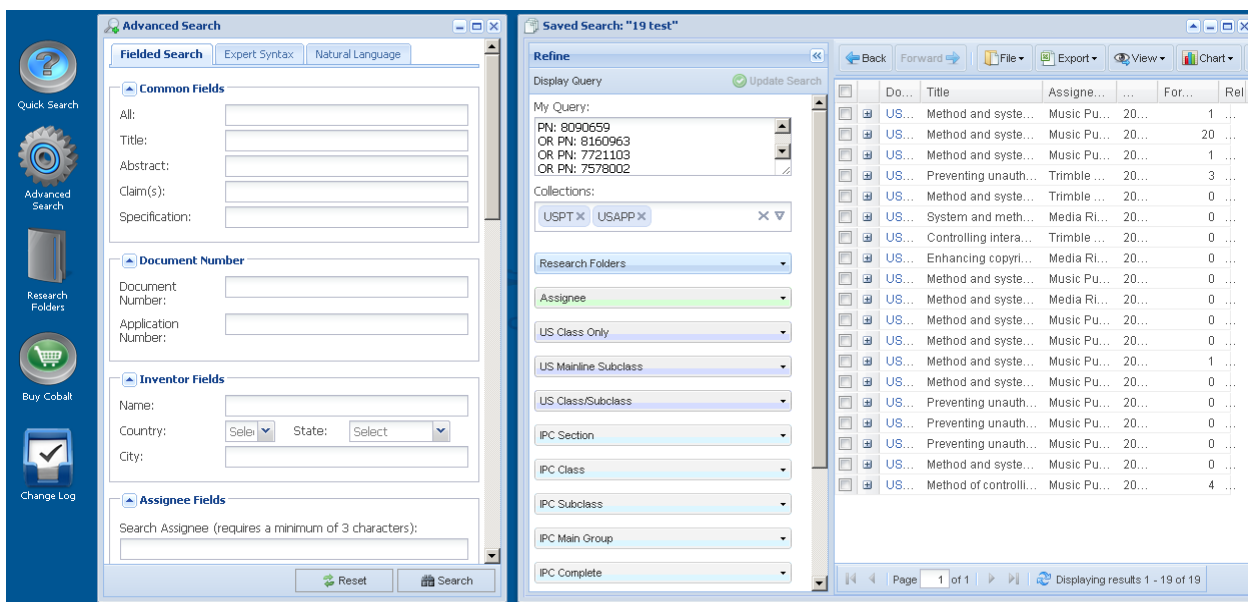
“Process Architect” је алат који нуди могућност креирања добро дизајнираних процеса, који омогућавају сарадњу унутар организације, усклађивање портфолија са мисијом компаније и максимизују повраћај инвестиција у истраживање и развој (енгл R&D) [13].

„Data Validation” служи за верификацију тачности патентних података кроз систем за управљање интелектуалним средствима [14]. Слика 4 показује пример визуализације обраде података помоћу „Data Validation” алата.



Слика 4: Пример визуализације „Data Validation”алата

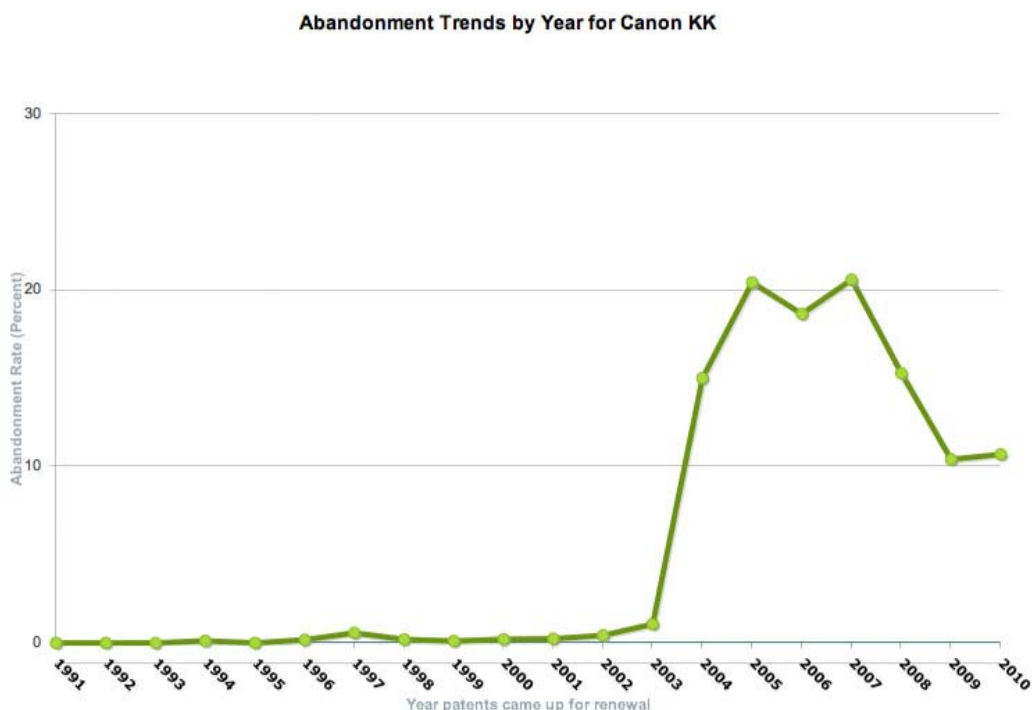
„CobaltIP“ [15] је алат заснован на коришћењу система прозора тако да корисник може отворити више прозора у исто време (Слика 5). Има преко 50 поља за претраживање података, која се могу применити као филтри. На пример, може брзо да сузи резултат претраге само на патенте које имају више власника (постоје су-власници додељени патентима).



Слика 5: Изглед „CobaltIP“ апликације са отворена два прозора

Овај алат представља решење намењено стручном истраживању. Садржи аутоматске функције које штеде време кориснику апликације као што је груписање докумената, која групише резултате претраге у функционалне концепте, тако да се могу лако открити важне теме које се налазе у подацима. Такође аутоматски детектује сличне документе, и

омогућава филтрирање тих резултата по одређеном власнику, класи, итд. Такође може да креира графиконе и визуализује резултате (поседује могућност приказивања десет различитих дијаграма). Слика 6 представља један пример визуализације резултата.



Слика 6: Визуализација резултата „CobaltIP“ апликације

„CobaltIP“ омогућава креирање графикона састављених од више серија података пружајући на тај начин могућност поређења тих серија, на пример, поређење патентног портфолија компаније *Apple* наспрам патентног портфолија компаније *Google* по класама САД на неколико нивоа детаља или по разним датумима. Омогућава, такође, да резултати анализе могу бити извезени у неке друге програме као што је *Excel*.

Оно што „CobaltIP“ издваја од других алата је што нуди нови приступ за проналажење сличних патената. Ова функција се зове ко-цитати (енгл *co-citations*), и аутоматски даје рангирану листу свих патената које деле сличне профиле цитираности са датим патентом.

3. Патенти

Шта су то патенти? Патент фактички представља ограничен монопол који држава одобрава одређеној особи или неком ентитету у замену за јавно обелодањивање проналаска. Прецизније речено то јесте монополско, територијално право које је при том и временски ограничено. Када једно лице постане носилац патента, оно има одређена права на одређеној територији (принцип територијалног важења закона), за одређено време (ограниченост правне заштите). Садржај права на основу патента одређују важећи законски прописи у свакој земљи. Својство носиоца права на патент доказује се помоћу патентне исправе, која представља формални доказ о постојању патента за одређени проналазак и права одређеног лица на тај патент. Патентну исправу издаје у свакој земљи одређени државни орган. Патент не условљава власника интелектуалне својине да упражњава дато решење, већ му даје право да спречи друге да користе, произведу и продају производ или услугу која користи патентирани проналазак.

Да би се заштитио патентом, изум мора на дан подношења пријаве за заштиту изума, односно на дан признатог права првенства пријаве:

- да буде нов,
- да има инвентивни ниво и
- да је индустријски примењив.

Патенти се признају за сваки проналазак, било да се ради о производима или поступцима, из разних области технике, под условом да су нови, да садрже инвентивни ниво и да су погодни за индустријску примену. Патент мора бити неочекиван или не очигледан. Иницијално, патентни испитивач одређује да ли је патент не очигледан тако што прегледа све релевантне патенте претходно објављене.

Проналазак је нов ако није обухваћен стањем технике. Сматра се да је проналазак обухваћен стањем технике ако је учињен доступним јавности пре датума подношења

пријаве изума, писаним или усменим путем. Стање технике чини садржај свих пријава проналазака које имају ранији датум подношења од пријаве која је у питању.

Проналазак има инвентивни ниво ако за стручњака из одговарајуће области не произилази, на очигледан начин, из стања технике.

Проналазак је индустријски примјенљив ако се предмет проналаска може произвести или употребити у било којој грани индустрије.

Исти проналазак може бити патентиран у више земаља на тај начин градећи тзв. фамилију патената. Ако је патент за одређени проналазак одобрен у једној земљи он има национални карактер, али његов утицај је присутан у целом свету јер та иста идеја може да буде патентирана било где.

3.1 Судски спорови

Патентне парнице су судски спорови који се покрећу у циљу утврђивања чињеничног стања да ли је одређени патент нарушен. Наиме, власник права над интелектуалном својином (проналазач, компанија или други ентитет), уколико сумња да је патент нарушен, тј. да неко други неовлаштено производи или продаје проналазак, покреће судски процес против дотичног лица или компаније. Такви судски спорови обично дуго трају и прилично су скупи. Случај се подноси оном суду који има јурисдикцију на територији где је повреда патентног права наступила. Сједињене Америчке Државе су једина земља у којој важи правило „први изумео“ (енгл. "*first to invent*") што значи прва особа која је изумела производ је власник патента, док у свим осталим земљама важи правило „први убележио“ (енгл. "*first to file*") што значи да је власник патента онај ко је први поднео патентну апликацију. Власник права над интелектуалном својином у том процесу настоји да докаже да је он заправо оригинални проналазач и да је патент нарушен од стране туженог. Тужени уколико жели да добије случај мора да докаже бар једно од следећег:

- да не нарушава наведени патент
- да је патент неприменљив
- да патент није валидан

Постоји неколико начина да се докаже да одређени патент није валидан. На пример, ако патент није нов или различит од свих осталих производа, метода или уређаја, онда се може сматрати да патент није валидан. Уколико се у току поступка постави питање очигледности патента, суд у том случају прегледа сву документацију како би се уверио да патент није очигледан узимајући у обзир датум проналаска. Уколико се покаже да је патент валидан, тада суд може да изда судски налог о забрани коришћења решења или

продавања предмета који нарушава патент. Додатно, власник над интелектуалним правом може бити обештећен и до три пута већим износом од онога колика је стварна штета која је проузрокована нарушавањем патентног права.

3.2 Издавање лиценци

Интелектуална својина је веома вредно право које има значајну улогу у пословању компанија. Највише директних користи од интелектуалне својине компанија има у случајевима судских спорова, лиценцирања, продаје па све до поступка банкротства. Власник патента има право на пренос права на друга лица путем наслеђивања патентног права или закључивањем лиценцих уговора. Власник права над интелектуалном својином (физичко или правно лице) може интелектуалну својину да пренесе, прода или поклони као сваку другу својину. Компанија, као носилац права над интелектуалном својином, чешће користи интелектуалну својину као економско право, него што је то случај са физичким лицем, односно индивидуалним проналазачем као носиоцем права.

Уговор о лиценци је уговор којим се обавезује давалац лиценце да стицаоцу лиценце уступи у целини или делимично право искориштавања изума, техничког знања и искуства, а стицалац лиценце обавезује се да му за то плати одређену накнаду. Уговор о лиценци мора се закључити у писменом облику. Лиценца за искориштавање патентираног изума, не може бити закључена на време дуже од трајања законске заштите тих права.

3.3 Вредновање патената

Право интелектуалне својине, као и неко друго финансијско добро, може бити искоришћено за повраћај новца зато што интелектуална својина, посебно индустријска својина (жигови, патенти, индустријски дизајн), постоји (региструје се и одржава) да би власнику права обезбедила вредне будуће токове новца. Власник права над интелектуалном својином мора да плаћа годишње таксе на име одржавања иновације, тј. да би њихови патенти били на снази. Износ накнаде расте сразмерно годинама патента тј. његовој старости.

Вредност патента се може дефинисати као разлика између вредности изума када изумитељ држи патентно право (монополско стање) и ако изумитељ не би имао патент на њега (конкурентно стање). Уопште гледано, вредност патената је врло тешко проценити. Мало патената чине заправо велик део њихове укупне вредности. На пример, недавна велика студија [16] показала је да само 7,2% од патената из узорка који је посматран су вредни више од 10 милиона евра (око 1% вреди више од 100 милиона евра), док око 68% од патената из истог узорка производе мање од 1 милион евра (око 8% вреди мање од

30.000 евра). Ови резултати се огледају у чињеници да многи патенти (изуми) се никада не искористе, а само неки од њих су преведени у комерцијално профитабилне иновације. Одлука да ли неки патент и како га користити, зависи од низа чинилаца. На пример, поседовање средстава који могу искориштавати патент је важно, као и стратешка одлука да се блокира конкуренција, да се побољша преговарачка моћ компаније у склапању лиценцих уговора, или како би се избегло да производи буду блокирани од стране конкурената. Споменута студија [16] идентификовала је шест различитих мотива за патентирање:

- пословна експлоатација иновација,
- лиценцирање,
- унакрсно лиценцирање,
- превенција од имитације,
- блокирање супарника, и
- углед.

Иста студија такође признаје да различите врсте послодаваца имају различите мотиве за патентирање. Комерцијална експлоатација патента је више важна за мала и средња предузећа, док унакрсно лиценцирање, превенција од имитације и блокирање конкуренције је важан разлог за патентирање за велике компаније. Међутим, један од најважнијих налаза је да мала предузећа и индивидуални проналазачи дају највећи допринос предузетним активностима на основу патената.

Квалитативне методе вредновања патената покушавају да рангирају патенте на основу јачине, ширине и правног статуса. Класичне квантитативне методе, односно традиционалне методе, се заснивају на:

- трошковном,
- тржишном или
- приносном приступу

Трошковни приступ је заснован на трошковима који су настали приликом добијања иновације заштићене патентом, било интерним развојем или екстерном аквизицијом. Овакав приступ је непристрасан и конзистентан. Поуздани индикатори вредности могу бити аквизициони трошкови патената.

Тржишни приступ користи упоредиве трансакције патената на тржишту. Ово је практичан приступ који обезбеђује примену актуелних цена за упоредива добра, тј. патенте. Проблем се јавља када се на тржишту не налазе патенти који би били упоредиви са патентом који се вреднује. Корелација између неких индикатора, као што је цитираност патената и вредности патената, је емпиријски добијена али комбинација тржишног

приступа и оваквих индикатора води до нових квалитетнијих начина вредновања патената уз обезбеђење објективне тржишне процене на вредност патента.

Приносни приступ покушава да израчуна садашњу вредност пројектовањем будућег тока раста прихода од предметног патента за време његовог економског животног века. Теоретски је супериоран у односу на друге наведене приступе јер се фокусира на будуће зараде или токове новца остварене на реализацији патента или производа на бази патента.

Традиционалне методе вредновања патената се заснивају на микро-економским подацима који укључују маркетиншке податке, продају на велико, приход остварен продајом производа и /или лиценцирањем патента. Поред ових метода, развијене су и статистичке методе вредновања патената. Тако се, на пример, статистичка метода „*Patent Value Predictor Model*“ заснива на макро економском моделу и приликом вредновања узима у обзир податке о укупној годишњој продаји производа који су покривени патентима и величину бруто домаћег производа (*Gross Domestic Product - GDP*).

Оно што додатно отежава одређивање вредности неког патента је то што у општем случају исти патент није једнако вредан за различите компаније. Разлози могу бити вишеструки, прво, неке компаније боље штите своја права интелектуалне својине од других компанија. Друго, вредност патента може зависити и од укупног портфолија компаније. Још један од чинилаца који утичу на вредност патента је животно век, или боље речено, рок трајања патента. Та вредност се мења током времена и такође зависи од степена заштите који је придружен датом патенту (на пример, да ли патент покрива технологију коју штити у уском или широком домену). Што патент више покрива технологију коју штити у широком домену (шири је), то може донети више профита његовом власнику.

Вредност самог патента може такође да зависи од укупне вредности патентног портфолија компаније и од улоге коју дотични патент има у том портфолију. Такође, уколико патент штити једно решење (метод или уређај) које је лако заобићи са релативно малом променом (тзв “лакоћа заобилажења патента”), то може да буде важан фактор у одређивању вредности патента. Још један од индикатора вредности патента су информације везане за обнављање патента као што је број земаља у којима је тражена заштита (тј, величина патентне фамилије). Цитираност патента је такође један од показатеља вредности патента. На пример, већа је вероватноћа да буду цитирани они патенти који су везани за неку важну иновацију. Вероватноћа литигације је такође један од параметара за одређивање вредности и она ће бити већа за вредније патенте.

Како патенти омогућавају власнику права над интелектуалном својином, законски забрањујући другима да користе дату иновацију, вредност датог патента се може утврдити

као разлика у повраћају новца која се стиче на основу иновације са и без постојања заштите интелектуалне својине. Откривање техничких информација може имати више користи за конкуренцију него за самог власника патентних права. У том случају можда је боље одлучити се за пословну тајну, пре него за патент.

Још 1996. године објављен је рад у којем је дискутовано о томе како библиографски подаци о патенту могу бити коришћени за бољу процену тј. унапређење израчунавања вредности иновација [17]. Представљена је једноставна шема заснована на броју година продужавања која може да унапреди излазну меру вредности иновације елиминишући и до половине шума у патентним рачуницама. Рад такође описује како помоћу наведених података може да се предвиди вредност власничких права (*proprietary rights*) која проистичу из патената и која могу даље бити кориштена за одређивање вредности патената.

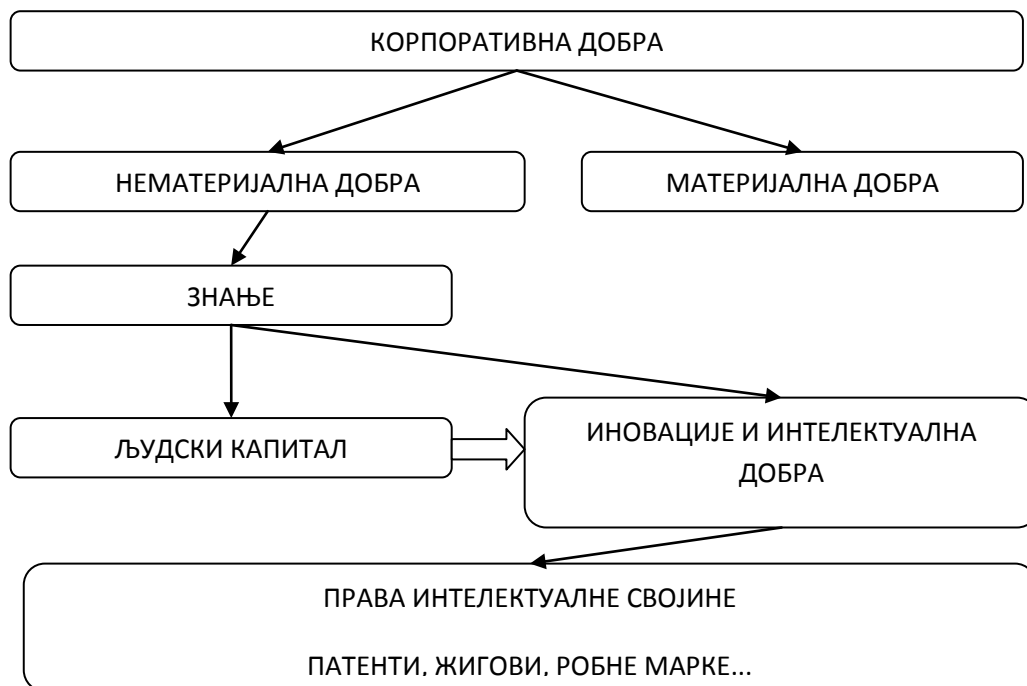
3.4 Допринос патената у укупној вредности компаније

Основни циљ финансирања интелектуалне својине је како идеју довести до производа или је превести у друго финансијско добро које доноси приход. Нематеријална добра, као и материјална, имају своју вредност и она мора бити на неки начин процењена. Да би се уопште дошло до процене вредности нематеријалног добра претходно се она морају идентификовати, морају имати дефинисаног власника и морају се регистровати. Нематеријално добро може имати и одређен животно век, на пример, патент важи двадесет година, а регистрација жига важи десет година, али се може обнављати неограничен број пута.

Поред проналазака, резултат проналазачког рада могу да буду и одређена знања и искуства која се примењују у индустријској или занатској производњи и привредном животу уопште. Уобичајено је и опште прихваћено, да се скуп ових знања и искустава заједничким именом назива “*know how*”. Израз “*know how*” се углавном не преводи на друге језике, већ се англосаксонски термин употребљава у говорном и писаном пословном и правном језику у целом свету. Национални прописи о индустријској својини и међународне конвенције из ове области не предвиђају правну заштиту за “*know how*”. Према томе, “*know how*” не представља правни монопол као остала права индустријске својине, али он има велику практичну вредност у поступку коришћења права индустријске својине као и осталих резултата проналазачког рада који не уживају правну заштиту. Онај који поседује “*know how*” поседује знање и искуство које му омогућава бржи и ефикаснији развој у односу на конкуренте који га не поседују. Према томе он ужива фактички монопол на тржишту. Битни елементи због којих се “*know how*” сматра фактичким правом

индустријске својине су његова тајност и преносивост, тј. способност да се знања и искуства која га сачињавају пренесе од једног корисника на другог. Предмет могу да буду знања и искуства потребна за индустријску примену једног патента, знања и искуства која представљају не пријављен (тајни) проналазак и сва остала техничка знања и искуства која се примењују у производњи.

У развијеној тржишној привреди стално је присутна утакмица између учесника на том тржишту и она даје импулс савременом технолошком развоју. Произвођачи непрестано усавршавају производњу у циљу одржања своје конкурентске способности. Они учесници у привредном животу који немају средстава или не желе да улажу средства у истраживање, настоје да прибаве потребан „*know how*” од оних који га већ имају. Због тога пренос знања, тј. трансфер технологије постаје исто толико значајан у свету колико и промет роба и услуга. Вредност знања је доказивана чињеницом да се сваке године милиони долара плаћају да би се стекло право његовог коришћења. „*Know how*” се може преносити путем уговора, најчешће уговора о лиценци, самостално, или заједно са неким заштићеним правом (патентом, моделом или жигом).



Слика 7: Састав корпоративних добара

Важност знања као покретачке силе у свету за економски раст знатно је порасла у последњих неколико деценија. Знање и производња нових идеја, тј. иновација и других интелектуалних добара, сада има централно место у развијеним компанијама, а нове технологије постају најважнији део у светској економији (Слика 7). У таквој поставци, власништво над патентима, тј. поседовање патентних права, постају један од најважнијих механизма за покретање пословања или за остваривање додатне добити. Улагање у

иновације постала је неминовност. Дакле, патенти су препознати као додатна вредност и све више добијају на важности.

3.5 Патенти као извори информација

Патенти су начини за заштиту интелектуалне својине, али и извршни извори информација. Патентни документи пружају релевантне техничке, правне и пословне информације које су јавно доступне и могу се искористити у више намена. Информацијама садржаним у патентима се, између осталог, може унапредити: разумевање тренутне позиције компаније на тржишту у смислу вредности њене интелектуалне својине, праћење вредности патената конкурентних компанија; прогнозирање технолошког развоја неке компаније (која је област активног истраживања), унапређење при доношењу одлука компаније где и када инвестирати, итд. Како су патентни портфолији предузећа постали изразито велики са по неколико хиљада, или чак неколико десетина хиљада патената, јасно је да се праћење и обрада овако великог скупа информација не може обављати без једног аутоматизованог система за обраду података.

Највећа информативна функција патената састоји се у њиховој могућности да открију и рашире знања о конкурентима тј. стању њихових производа и процеса. Такође омогућава компанији да развије програм лиценцирања сопствених патената, што може постати значајан нови извор прихода. Приликом анализирања патентних портфолија при конструисању било каквих закључака или тврдњи, треба узети у обзир и величину предузећа. Наиме, недавна студија [18] показује да мања предузећа са већом вероватноћом уступају своје патенте на коришћење или дају лиценце за своје технологије у односу на већа предузећа. Међутим, исто истраживање показује да је мања вероватноћа да ће мала и средња предузећа користити туђе патенте као извор информација за своје иновационе активности у односу на веће компаније. Постоје барем два узрока овог проблема: прво, скуп и компликован поступак за коришћење патентних база података и алати који нису приступачни за мала и средња предузећа, а други, недостатак разумевања вредности информација које су садржане у патентној документацији.

Патентна апликација је заправо први документ у којем се описује и открива неки изум. Постоје тврдње да се чак око 80% свих научних и техничких информација могу пронаћи само у патентним документима [19].

3.6 Структура патента

Патентни документ има прилично структуриран формат који се објављује у складу са међународно прихваћеним стандардима. Патентни подаци приказани су на систематски начин, као комбинација структурираних и не структурираних података.

Патентни документ почиње с библиографским подацима патентираног изума (Слика 8). Осим тога, на првој страници су и наслов патента и сажетак којем може бити придодат и илустративни цртеж изума. Након насловне странице, налази се главно део патентног документа. У већини случајева, главни део патентног документа укључује стандардне садржаје о изуму по следећем редоследу:

- идентификација технологије или техничког подручја на које се изум односи
- процена претходног стање технике у техничком подручју тзв. позадина изума
- техничка расправа о проблему или проблемима које изум намерава решити
- опис изума довољно детаљан у односу на један или више примера употребе, како би квалификоване особе биле у могућности да реализују решење из патента
- цртеж(и) у вези с примером употребе, у сврху бољег разумевања изума наведеног у опису.



US006870886B2

(12) **United States Patent**
Challapali et al.

(10) **Patent No.:** US 6,870,886 B2
(45) **Date of Patent:** *Mar. 22, 2005

(54) **METHOD AND APPARATUS FOR TRANSCODING A DIGITALLY COMPRESSED HIGH DEFINITION TELEVISION BITSTREAM TO A STANDARD DEFINITION TELEVISION BITSTREAM**

(75) Inventors: **Kiran S. Challapali**, Ossining, NY (US); **We dong Mao**, Briarcliff Manor, NY (US)

(73) Assignee: **Koninklijke Philips Electronics N.V.**, Eindhoven (NL)

(*) Notice: This patent issued on a continued prosecution application filed under 37 CFR 1.53(d), and is subject to the twenty year patent term provisions of 35 U.S.C. 154(a)(2).

Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 900 days.

(21) Appl. No.: **08/167,394**

(22) Filed: **Dec. 15, 1993**

(65) **Prior Publication Data**

US 2002/0001343 A1 Jan. 3, 2002

(51) **Int. Cl.**⁷ **H04N 7/12**

(52) **U.S. Cl.** **375/240.24**

(58) **Field of Search** 348/441, 445, 348/458, 389, 388, 415-417, 426, 427, 431, 430, 409-413, 420, 419, 469, 500, 699, 390; 375/245, 240.03, 240.12-240.16, 240.18, 240.2, 240.21, 240.24; H04N 7/12, 7/01

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,475,213 A	*	10/1984	Medaugh	375/245
4,860,313 A	*	8/1989	Shapiro	348/415
4,866,519 A	*	9/1989	Lucas et al.	348/469
4,893,123 A	*	1/1990	Boisson	348/419
5,055,927 A	*	10/1991	Keesen et al.	348/389
5,243,428 A		9/1993	Challapali et al.	358/167
5,253,056 A	*	10/1993	Puri et al.	348/415
5,262,854 A	*	11/1993	Ng	348/416
5,270,813 A	*	12/1993	Puri et al.	348/384
5,589,993 A	*	12/1996	Naimpally	348/390
5,764,298 A	*	6/1998	Morrison	348/500

OTHER PUBLICATIONS

Le Gall, "MPEG: A Video Compression Standard for Multimedia Applications", Communications of the ACM, vol. 34, No. 4, Apr. 1991, pp. 46-58.

Girons et al., "The Digital Simulcast AD-HDTV Coding System", IEEE Transactions on Consumer Electronics, vol. 38, No. 4, Nov. 1992, pp. 778-783.

"Information Technology—Generic Coding of Moving Pictures and Associated Audio", Recommendation H.262, Nov. 1993, ISO/IEC Committee Draft.

"Test Model 5: Draft—Test Model Editing Committee", ISO/IEC, Apr. 1993.

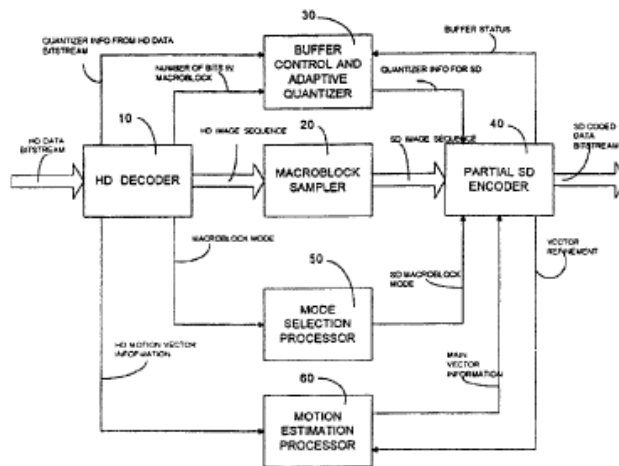
* cited by examiner

Primary Examiner—Richard Lee

(57) **ABSTRACT**

A method and apparatus for transcoding macroblocks of a high definition television signal to co-sited macroblocks of a standard definition television signal wherein SD macroblocks are derived directly from co-sited HD macroblocks, i.e. without having to compute the SD macroblock information from its corresponding SD image sequence.

16 Claims, 3 Drawing Sheets



Слика 8 Изглед прве странице патентног документа

Патентне информације су објављене или у патентним документима или произлазе из анализе статистичких података из патентних пријава. Технички подаци су изведени из описа и из цртежа изума; правни из патентних захтева (тврдњи) и из његовог правног статуса; пословни из референтних података о проналазачу, датуму подношења патентне пријаве, земљи порекла, итд.; и јавни из анализе трендова у подношењу патентних пријава.

Структура и садржај патентног документа нуде пуни спектар могућности за коришћење патентних информација у области технолошког менаџмента и пословне интелигенције. Патентне информације могу помоћи корисницима: да избегну додатно улагање у истраживање и развој, да избегну нарушавање патентних права других проналазача, да би се проценила вредност сопствених или патената других проналазача; да се истраже технологије из патентних пријава које никада нису одобрене, које не вреде у неким земљама, или које више нису на снази; да се прикупи знање везано за иновативне активности и предвиде истраживачка кретања пословних конкурената, да се унапреди планирање и доношење пословних одлука, као што су лиценцирање, склапање партнерства или спајања и преузимања, те да се идентификују кључни трендови у појединим техничким подручјима.

Заинтересовани за патентне информације су у највећем броју научници и проналазачи, предузетници и комерцијалне компаније. Да би били у стању да донесу праву одлуку о коришћењу и управљању патентима они морају разумети све предности, слабости, могућности или претње које се односе на управљане конкретним патентним средствима. Поред тога, управљање патентним портфолијом подразумева разумевање тренутне позиције компаније на тржишту, праћење конкуренције; прогнозирање технолошког развоја, доношење одлука унутар компаније: где/када инвестирати, и одабирање адекватне одбрамбене тактике.

Да би у потпуности искористили потенцијал патентних докумената и извукли драгоцене садржаје из њих, неопходно је анализирати структуриране као и не структуриране податке [20]. То значи да треба прикупити и анализирати основне статистичке податке из патентних докумената, на основу информација са насловне стране (библиографски подаци, подаци у структурираној форми), али и анализирати слободан текст (не структурирани текстуални подаци) у телу документа како би се обезбедила смислена повратна информација корисницима. Табела 1 показује однос структурираног и не структурираног текста у одређеним деловима патентног документа.

Део патентног документа	Где се налази	Садржај	Тип	Намена
Наслов и сажетак	Насловна страна	Сажетак о технологији проналаска	Слободан текст	Архивирање
Библиографски подаци	Насловна страна	Број документа, датуме подношења пријаве и објављивања, имена проналазача, итд.	Структуриран	Пословне информације од јавног значаја
Објашњење	Главни текст	Детаљно техничко објашњење проналаска	Слободан текст	Техничка и правна
Захтеви	Главни текст	Дефинише тачно област која се штити проналаском	Слободан текст	Правна

Табела 1 Типови података у појединим деловима патентног документа

4. Развој техничког решења

Проблем се своди на прикупљање патентних докумената и аутоматску обраду текста како би се ефикасно вредновао портфолио неке компаније. У овом поглављу биће изложена методологија рада којом се жели доћи до техничког решења постављеног проблема. Предложени метод биће верификован коришћењем стварних патентних портфолија неколико компанија које су такође биле предмет разматрања патентних експерата. Очекује се да ће ова метода допринети смањивању времена које је потребно експертима да процене патентни портфолио компаније и допринети ефикаснијем и бољем руковању патентним портфолијом.

Методологија развоја техничког решења за аутоматску анализу патентних докумената је развијена и може се представити следећим фазама:

1. Прикупљање патентних докумената и складиштење одговарајућих поља патентног документа у базу података
2. Обрада података из базе података:
 - а) Мерење фреквенције кључних речи у сваком патенту и у свим патентима из одабраног скупа патената
 - б) Редукција матрице више-димензионог простора у простор две или три димензије
 - с) Класификација
3. Приказ резултата и генерисање извештаја

4.1 Прикупљање и складиштења патентних докумената

Прикупљање патентних докумената и њихово складиштење на рачунару због даље обраде је главни задатак једног дела програмске подршке – интернет робота. Документи који су предмет обраде су заведени и јавно доступни на интернет сајту патентне

организације: Америчког завода за patente и трговачке марке - *United States Patent and Trademark Office (USPTO)*. Поред прикупљања веома важан део је парсирање и издвајање библиографских података, као што су: проналазач, подносилац, датум пријаве патента, међународни класификациони код, цитиране публикације, статус патента итд. Издвојени подаци се затим складиште у базу података.

4.2 Обрада патентних докумената

У обраду патентних докумената спада парсирање и анализа патентног текста. Анализа патентног текста изискује употребу метода за тзв. „откривање знања“ које анализирају текст из више перспектива и издвајају из текста корисну информацију. Укратко, методе се користе за обраду података и издвајање информација. Сирови подаци (не структурирани текст) се трансформишу у структуриране податке. У циљу издвајања корисне информације из збирке докумената, алгоритми идентификују и издвајају кључне речи и истражују обрасце међу њима или издвајају кључне речи. На основу издвојених кључних речи формира се високо димензиона матрица над којом се затим ради редукција димензионалности и класификација. Дакле основни кораци су:

- (i) Парсирање и анализа патентног текста, укључујући и лексичку анализу
- (ii) Редукција матрице више-димензионог простора у простор две или три димензије, тако да се одржи највећа могућа сличност структури оригинала
- (iii) Аутоматска класификација патената по сличности контекста.

4.2.1 Анализа патентног текста

Анализа патентног текста укључује пред-процесирање патентних докумената у смислу провере редувантности података, статистичке анализе, парсирања и структурирања текстуалних и нумеричких података. Поред тога анализа патентног текста укључује и лексичку анализу текста, нпр. налажењем учестаности појаве речи у сваком од докумената, као и кумулативне фреквенције појаве у свим документима. Једна од таквих метода је и позната метода: „*Text Frequency – Inverse Document Frequency (TF-IDF)*” која се често користи као шема за издвајање пондерисаних кључних речи [21]. Ова метода је кориштена и у овом техничком решењу.

4.2.2 Редукција података

Редукција података је базирана на истраживању сличности или разлика у подацима које обрађује. Као резултат ове методе високо-димензиони подаци садржани у патентним документима се трансформишу у простор много ниже димензионалности (2Д) тако да при

том одржава највећу могућу сличну структуру оригиналу. Редукција података садржана у патентним документима обухвата следеће кораке:

- (i) Формирање лексичког модела.
- (ii) Примена постојеће методе за редукцију података која је погодна за обраду текста

Једна од најпознатијих метода за редукцију података је класично мултидимензионо скалирање - *Multi-Dimensional Scaling (MDS)* метода [22] која ће бити кориштена и у овом техничком решењу.

4.2.3 Класификација

Аутоматска класификација патената сличних по контексту се заснива на издвајању и одређивању важности свих речи дужих од три карактера из анализираног скупа патената. Важност анализираних речи се изражава кроз тежински фактор речи и она представља нумеричку вредност која зависи од фреквенције (броја понављања) анализираних речи у односу на све остале речи скупа патената. Филтрирањем речи на основу израчунатог тежинског фактора над целим скупом патената добија се могућност приказивања базирано на контексту у матричном облику са нумеричким вредностима. Коришћењем алгорита за редукцију димензија матрица као и алгорита за класификацију података матрица добија се могућност одређивања кластера над скупом патената као и одређивање њихових кључних речи.

Циљ кластер анализе је да формира групе објеката који су међусобно најсличнији у односу на објекте из других група. Сличност је обично представљена као нека врста дистанце, а најпознатије су Еуклидове дистанце. Еуклидова дистанца је једноставна примена питагорине теореме. Она је оптималан избор када је простор метрички, континуалан, односно квантитативан и када променљиве нису у корелацији. Постоји велики број различитих показатеља дистанци намењених различитим типовима података. Показатељи корелације, као мера сличности, такође се могу користити као инверзна мера удаљености – дистанце. Показатељи асоцијације бинарних променљивих такође могу послужити као инверзни показатељи дистанце.

Постоје разне методе кластер анализе. Могу се поделити у две основне групе: хијерархијске и не хијерархијске.

Хијерархијске методе су по правилу секвенцијалне и не преклапајуће. Најчешће су то агломеративне методе јер стартују од одређеног броја објеката и секвенцијално их разврставају у све веће кластере, а ређе су дивизивне, када полазе од једног кластера и разбијају га на ситније. Типично, у сваком кораку се формира један кластер спајањем два

најближа кластера из претходног корака. Који кластери ће бити кандидати за спајање одређује се на основу већег броја различитих техника. Хијерархијске методе имају недостатак у томе што је матрица дистанци димензија $N \times N$. Зато када је N велико, често уопште није могуће извести кластер анализу. Не-хијерархијске методе не полазе од матрице дистанци, него објекте сврставају у кластере чији је број, или центроид, унапред задат.

Поступак не-хијерархијске класификације почиње поделом иницијалног скупа објеката у изабрани број група или према иницијално одређеним центроидима за сваку групу. Затим се одреде одстојања између сваког објекта и сваке групе (иницијалног центроида). На основу одстојања објекти се лоцирају у њима најближе групе. По придруживању објекта групи израчунава се центроид групе из које је објекат прешао и групе којој се објекат придружио. Поступак се понавља за сваки објекат тако што се поново рачуна његово одстојање од центроида група и врши прерасподела између група. Најпознатија не-хијерархијска метода је „к-просека“ (енгл *k-means*). Почиње се од k иницијалних кластера. Објекти се разврставају у онај кластер чији им је центроид најближи. Иницијални кластери се могу одредити арбитарно или на основу неког претходног груписања. Постоје и побољшања ове методе као што је техника покретних облака (енгл *running means*).

Четири основна алгоритма за груписање која су анализирана у овом раду су: „*k-means*“, „*fuzzy c-means*“, „*neural gas*“ и „*re-organizing neural network –RONN*“ [23].

„*K-means*“ метода представља не-хијерархијски метод груписања. Заснована је на покушајима да се открије K оптималних кластера методом итеративних прегруписања на основу оптимизацијске функције. Посматрани скуп се подели у K кластера у коме свака опсервација припада кластеру који има најближу средину. То значи да користи центар као представник кластера. Подразумева следеће кораке:

- 1) одреди се број класа у које ће се сврстати полазни скуп ентитета;
- 2) бира се почетни центроид;
- 3) примењује се општи поступак за било коју не-хијерархијску методу класификације, а критеријум заустављања је:
 - понављање сукцесивних решења;
 - достизање максималног броја итерација;
 - постизање задате хомогености.

„*Fuzzy c-means Clustering - FCM*“ је метода класификација која као резултат даје једну или више континуалних променљивих које представљају вероватноћу припадности неком кластеру. То је основна метода фази груписања која је веома слична методи „*K-*

means“, разлика је да сваки објекат има степен припадности кластерима. Тако, објекти на ивицама кластера су у кластеру у нешто мањој мери од објеката унутар кластера.

„*Re-Organising Neural Network – RONN*“ је итеративан поступак учења, која обавља итеративни прилагођавања координата чвора, у маниру K -просек алгоритма, све док се чворове релативно не стабилизују на својим тренутним позицијама, а истовремено помера оне чворове за који се испостави да су мртви чворови, на боље позиције.

„*Neural Gas*“ је техника учења без надзора која омогућава јединствени пласман репрезентативних прототипова у векторском простору. Овај алгоритам одређује прототипове на такав начин да Еуклидова дистанца између вектора података и вектора прототипова је минимална. То је релативно једноставан алгоритам за проналажење оптималних представника података и представља снажну алтернативу за „*K-means*“ начин груписања. Широко се користи у областима где је потребна компресија података, попут обраде слике или препознавање облика.

Најчешћи проблеми кластер анализе су постојање великог броја показатеља сличности и великог броја метода. Намеће се проблем како одредити дистанцу и методу, као и како одредити број кластера, тј. како одредити квалитет решења. Избор дистанци зависи од: типа података, врсте проблема и врсте простора. Ако су подаци континуални, Еуклидова метрика је обично одговарајућа, међутим за не континуалне типове података погодно је користити друге метрике. Када је у питању избор методе за класификацију, постоје многа истраживања и симулације којима су испитиване различите методе кластер анализе. Неке методе су осетљиве на екстреме, а неке нису, неке теже да дају кластере једнаке величине, а неке не. Неке методе теже ка малим компактним кластерима, а друге ка великим, разуђеним. Нажалост, нема једноставног и универзалног упутства. Одређивање броја кластера је проблем сличан одређивању броја фактора у факторској анализи, али је још теже решив. У различитим софтверским имплементацијама постоје различити показатељи који требају да укажу које решење је најбоље, али ниједан није универзалан и опште прихваћен. Данас се провера кластер анализе и оцена квалитета коначног решења најчешће обавља путем каноничке дискриминационе анализе. Као групишућа променљива се користи “кластер променљива”. Осим дискриминационе анализе, провера се може вршити анализом варијансе, тзв. t -тестом, табелама контингенције и сл.

Кластер валидност представља мерење колико је добро груписање у односу на друга груписања креирана од стране других алгоритама, или од стране истих алгоритама коришћењем других вредности параметара. Кластер валидација је веома важно питање у анализи груписања јер резултат груписања треба да буде вреднован у већини апликација.

У већини алгоритама, број кластера је постављен као улазни параметар корисника. Постоји много приступа да се пронађе најбољи број кластера.

Једна од квантитативних мера за оцену груписања је и *Davies-Bouldin Index (DB Index)*. То је интерна евалуациона шема, где се валидација колико добро је груписање урађено постиже коришћењем карактеристика својствених улазним подацима. Показује раздаљину између објекта који су ван кластера и објеката који се налазе унутар кластера.

$$DB = \frac{1}{n} \sum_{i=1}^n \max \left(\frac{S_n(Q_i) + S_n(Q_j)}{s(Q_i, Q_j)} \right) \quad (1)$$

где је n број кластера, S_n средња дистанца свих објеката у кластеру од центра кластера, $S(Q_i, Q_j)$ дистанца између кластер центара. Однос је мала вредност уколико су кластери компактни и удаљени један од другог. Тако да, *Davies-Bouldin index* ће имати малу вредност за добро груписање.

Још једна од квантитативних метода за оцену груписања је и бројање патената који су ван заједнице (енгл *outliers*).

$$\text{dist}((x_i, y_i), C_j) - S_n > N * \text{std}(\text{dist}) \quad (2)$$

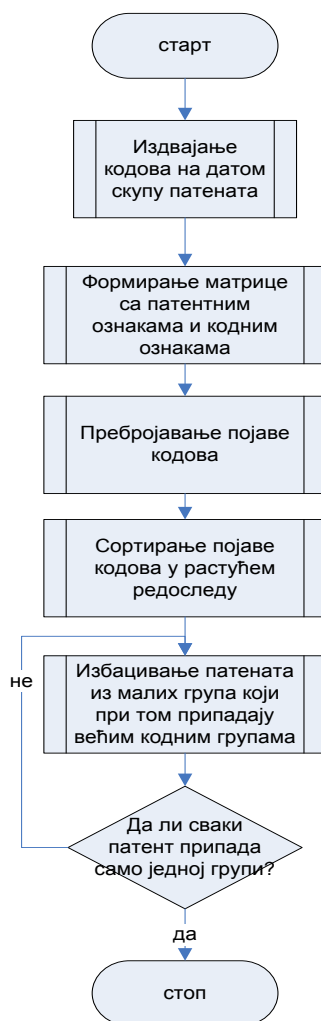
где је, dist Еуклидова дистанца, C_j центроид кластера, S_n средња дистанца свих објеката у кластеру од центра кластера, N је вредност којом се дефинише област око центроида кластера која се сматра пожељном (типично се узима 1.5 за строже, или 3 за блаже услове).

У настојању да се унапреди процес класификације патената у овом истраживању се покушало да се у тај процес укључе и међународни *IPC* кодови. У ту сврху развијена су два алгорита за класификацију на бази кодова.

4.2.3.1 Метода за класификацију на бази кодова

Како би се дошло до што бољих резултата за класификацију патентних докумената развијен је властити алгоритам који је базиран на анализи међународних класификационих кодова који су придружени сваком патенту. Алгоритам се састоји од следећих корака:

- издвоје се сви кодови на датом скупу патената
- израчуна се број појављивања сваког кода
- врши се сажимање патената по кодовима тако што се тежи што већим групама



Слика 9 Алгоритам методе за класификацију

Пример

Претпоставимо да имамо скуп од три патената у којима се укупно појављују 4 класификациона кода, посматрајући само прва четири карактера. На основу предложеног алгоритма формира се табела где прва колона садржи каталожку ознаку патената, а преостале колоне представљају кодове који се појављују у патентима (Табела 2).

PID	H03M	H04B	G06K	G06T
US123456	1	0	1	1
US415543	0	1	0	1
US789012	0	0	1	1

Табела 2 Пример сортирања патената по кодовима

Затим се израчуна број појављивања сваког кода. Добијена листа се сортира у растућем редоследу (од најмање ка највећој вредности) (Табела 3).

	H03M	H04B	G06K	G06T
број појављивања:	1	1	2	3

Табела 3 Пример пребројавања

Листа се затим анализира тако што за сваку колону која има мали број патената избацује се патент који постоји у некој од других бројнијих колона, све до најбројније колоне како не би било појављивања патената у више колона. Сажимањем патената у групе избацили би се патенти *US123456* из прве колоне „*H03M*” (јер постоји у бројнијим колонама „*G06K*” и „*G06T*”) и *US415543* из друге колоне „*H04B*” (јер постоји у бројнијој колони „*G06T*”), затим патент *US123456* из колоне „*G06K*” и патент *US789012* из колоне „*G06K*”. Као резултат добија се број патената по некој од ознака (Табела 4).

IPC1	PID	COUNT(ipc1, pid)
H03M	US123456	1
H04B	US415543	1
G06K	US123456	2
G06K	US789012	2
G06T	US123456	3
G06T	US415543	3
G06T	US789012	3

Табела 4 Резултат алгоритма

4.2.3.2 Метода за класификацију на бази кодова и технологија

Као унапређење претходне методе развијена је и њена напреднија верзија која би на основу класификације на бази кодова требало да на аутоматизован начин идентификује и технолошку област којој одређена група патената припада.

Класификација патената обавља се на основу претходно дефинисаних технолошких група. Технолошке групе формирају експерти за дату област, они свакој групи додељују претходно анализирани патенте. Сваку технолошку групу описују *IPC* кодови патената који су јој додељени. Ради лакше анализе предвиђено је да постоје главне технолошке групе, од којих свака може имати више подгрупа. Патенти се искључиво додељују подгрупама, тако да је неопходно да свака главна група има бар једну под групу. Приликом формирања групе обавља се статистика обрада *IPC* кодова на следећи начин:

1. За сваки *IPC* код чува се његова фреквенција у оквиру технологије
2. За сваку технологију се чува:
 - a. укупан број патената на основу којих је технологија формирана

- b. Укупан број јединствених *IPC* кодова
- c. Фреквенција заступљености *IPC* кодова у патентима који припадају датој подгрупи

Ове вредности ће се користити приликом класификације нових патената. На основу претходно дефинисаних технолошких група – могуће је класификовати нове патенте по технологијама којима припадају на основу својих *IPC* кодова. Аутоматска класификација се обавља на следећи начин:

1. За сваки патент се узимају *IPC* кодови
2. За сваку под групу се преузимају статистички подаци уколико се неки од *IPC* кодова у тој групи појављује
3. Ако имамо једну технолошку под групу која има највећи број *IPC* кодова – тој групи се додељује патент
4. Ако имамо више подгрупа које имају исти број *IPC* кодова, анализа се обавља на следећи начин:

Рачуна се следећи коефицијент за сваку од технолошких група

$$K = \sum_i f_{IPC_i} \frac{K}{N_{pat}} \quad (3)$$

где је N_{pat} број патената у технологији, $K = \sum f_{IPC}$ сума фреквенције понављања свих *IPC* кодова у оквиру једне технологије, а f_{IPC} број појављивања датог *IPC* кода у технологији. Она група која има највећи коефицијент представља технолошку групу којој дати патент припада.

Предност ове методе у односу на претходну је што може да идентификује технологију којој одређена група патената припада. Међутим, мана јој је што мора постојати довољно добар предефинисан скуп већ класификованих патената како би се алгоритам на довољно добар начин обучио. Једно решење је да се метода примењује на специјализованом домену технологија који је могуће довољно добро дефинисати.

4.3 Визуализација

Приказ резултата своди се на визуализацију кластера добијених из редукованог простора патената у жељеном контексту (технологије, компаније, цитираност, итд). [24]. Програмска подршка треба да омогући визуализацију података са неколико аспекта као што су: технолошки кластери, груписање на основу власника патентног права, расподелу патената у времену или по класификационим кодовима. Детаљнији опис програмске подршке и њених могућности дат је у поглављу 5.

4.4 Скуп Података

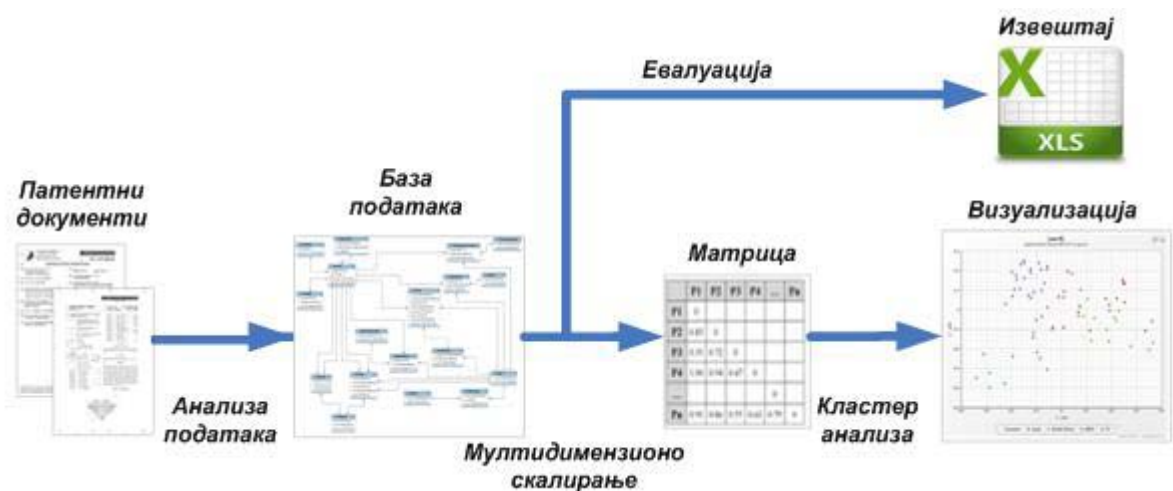
Иако су вештачка интелигенција и машинско учење значајно побољшани током последњих неколико година, патентна анализа и груписање патената уз помоћ стручњака остао је ипак најбезбеднији и најпрецизнији начин. Због тога, ће се резултати програмске подршке упоређивати са резултатима експерата. У овом раду користиће се скуп од око 1000 патената који ће послужити као улазни подаци за тестирање техничког решења. Да би се упоредиле перформансе одређених делова техничког решења упоредиће се резултати програмске подршке са резултатима стручњака који су дали своја мишљења о вредности патената и креираним кластерима. У сврху тестирања опште познатих техника груписања и верификације резултата (поређењем са референтним резултатима експерата), генерисаће се код у програмском пакету „*Matlab*“. Приликом верификације користиће се две методе поређења: математичко и визуелно поређење [25].

5. Програмска подршка - ПСАЛМ

Један од главних резултата овог научно-истраживачког рада је интернет оријентисана програмска подршка која служи за прикупљање и обраду патентних докумената. У то конкретно спада:

- повлачење и пред-процесирање патентних докумената (провера редуваности, статистичка анализа, парсирање и структурирање текстуалних и нумеричких података).
- редуција података садржаних у тексту патентних докумената заснована на лексичком моделу.
- анализа редуваног простора на кластере података и дво-димензиону (2Д) визуализацију група тачака по изабраном критеријуму: приказ технологија, патената, компанија, релација цитираности итд.
- генерисања извештаја.

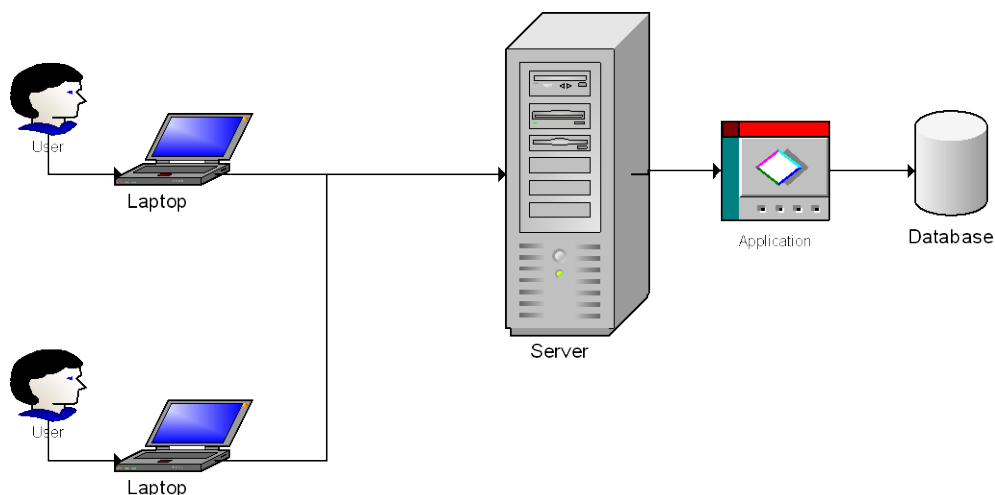
Слика 10 показује генерални начин функционисања програмске подршке.



Слика 10 Основни модел програмске подршке

Програмска подршка за аутоматску анализу и обраду патентних докумената названа ПСАЛМ (*PSALM – Patent Search and Analysis for Landscaping and Management*) развијена је у сарадњи научника са одсека за рачуарску технику и рачуарске комуникације факултета техничких наука, универзитета у Новом Саду и стручњака из института за рачуарску технику и рачуарске комуникације: РТ-РК, такође из Новог Сада. Алат је развијен у *Java* и *PHP* програмском језику и заснива се на бази података (енгл *MySQL*). ПСАЛМ проналази патентне документе на јавно доступним базама података, прикупља их и анализира библиографске параметре патената, али такође и текст патената. Више-димензиони подаци садржани у патентним документима се сажимају у простор знатно ниже димензије (2Д), групишу се и затим визуално приказују.

Програмска подршка је структурирана и може опслуживати више умрежених корисника у исто време у сервер - клијент техници (Слика 11). Корисничка спрега је изграђена коришћењем *PHP*, *HTML* и *JavaScript* програмског језика, *jQuery JavaScript* библиотека, као и *DataTables* и *HighCharts* библиотека за приказ резултата обраде података.



Слика 11 Генерална поставка система

ПСАЛМ је алат развијен за анализу већег броја патената. Цео систем је направљен и ради на бази појединачних случајева, у коме сваки случај се састоји од групе одабраних патената на бази критеријума које дефинише корисник. Унапред дефинисани критеријуми за формирање случаја су базирани на: правном заступнику (опуномоћенику), *IPC* кодовима и цитираним и патентима који цитирају дати патент. Осим наведених критеријума, корисник може креирати неограничен број критеријума за груписање кључних речи патента, сортирање и тако даље. Сваки случај је након креирања непроменљив. Међутим, могуће је створити нови случај са другим скупом патената комбиновањем постојећих случајева.

CASE select Progress Report Charts ljuba is logged in | logout

powered by Patent Core Team New CASE Select Patents by PID

Patent List

CASE ID: 5 CASE Title: Test case, patents loaded from CSV file name: randompatents.csv CASE Description: CSV file name: randompatents.csv

PID	Title	Assignee
US5323396	Digital transmission system, transmitter and receiver for use in the transmission system	U.S. Philips Corporation (New York, NY)
US5544247	Transmission and reception of a first and a second main signal component	U.S. Philips Corporation (New York, NY)
US4972484	Method of transmitting or storing masked sub-band coded audio signals	Bayerische Rundfunkwerbung GmbH (Munich, DE)
US4833543	Image processing system and phaselocked loop used therein	Alcatel N.V. (Amsterdam, NL)
US4970590	System and device for package multiplexing in transmission of many data flows generated by a sole algorithm	Telettra - Telefonica Elettronica e Radio (IT)
PID: US4970590 Title: System and device for package multiplexing in transmission of many data flows generated by a sole algorithm Abstract: A system for transmission of signals coming from a source and processed by an algorithm that, to minimize data (values+parameters), generates values that are encoded, for example, with variable length. The data flows generated by the algorithm are each ordered in packages with the addition of information of source, of frame and of management before being multiplexed.		
US5365272	Method for formatting compressed video data into transport cells	General Electric Company (Princeton, NJ)
US5453790	Video decoder having asynchronous operation with respect to a video display	Alcatel N.V. (Amsterdam, NL)
US5291284	Predictive coding and decoding with error drift reduction	British Telecommunications (London)
US4982270	Video data transmitting system	Canon Kabushiki Kaisha (Tokyo, JP)
US5068724	Adaptive motion compensation for digital television	General Instrument Corporation (Hartboro, PA)

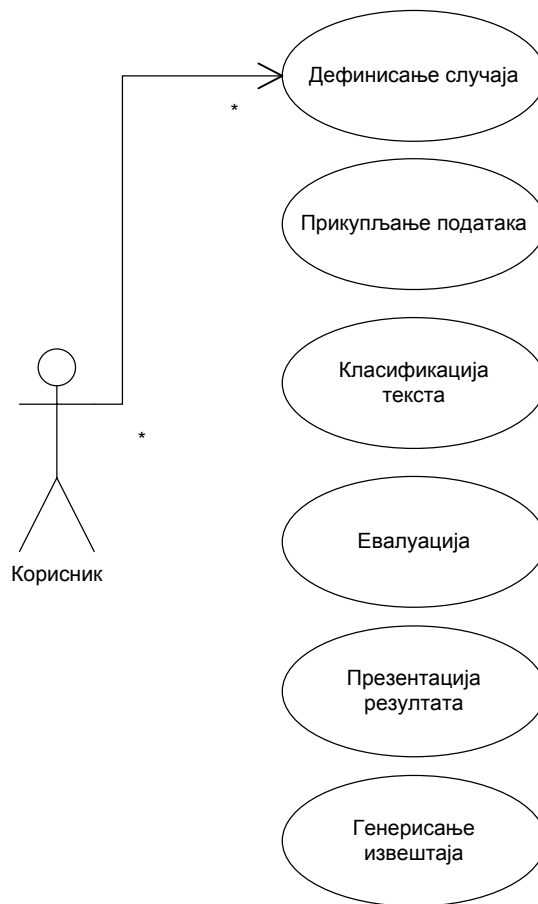
Showing 1 to 10 of 143 entries First Previous 1 2 3 4 5 Next Last

Слика 12 Пример изгледа главног прозора

Програмска подршка се састоји од следећих функционалних модула:

- интернет робота,
- класификације текста у виду кластера,
- сажимања вишедимензионалног простора,
- визуализације,
- анализе међународних патентних кодова и класификације,
- издвајање цитираних и патената који цитирају дати патент,
- модула за снимање података у излазну датотеку и
- модула за евалуацију (вредновање) патента.

Слика 13 приказује један случај употребе програмске подршке, тј. интеракцију између модула и корисника.



Слика 13 Случај употребе

Псалам-ом се анализирају и структурирани и не структурирани подаци о патентима.

- Алат прикупља патентне податке из јавно доступне базе података „USPTO“ и анализира библиографске параметре патената (као што су: наслов, проналазач, подносилац пријаве, датум пријаве, првобитни датум, земља издавања, првобитни број, првобитна земља, референце цитираних патената, референце патената који цитирају дати патент, сажетак, међународне класификације патената) [26]
- Анализира текст патената помоћу тежинске шеме фреквенције појаве речи (*Text Frequency - Inverse Document Frequency TF-IDF*) [27].
- Више-димензиони подаци садржани у патентним документима се претварају у знатно нижу димензију простора (2Д), задржавајући сличну структуру оригиналном простору, коришћењем шеме вишедимензионалних скала (*Multidimensional Scaling- MDS*).
- Редуковани простор патентних података је груписан помоћу технике не надзиране кластеризације како би се груписала задана неозначена колекција патената у смислене кластере. Овај приступ омогућује да се корисне информације извуку из патената кроз идентификацију и анализирање кључних речи и кључних фраза из текстуалних података патената.

- ПСАЛМ омогућује визуализација високо и ниско димензионих података. Омогућује визуализацију високо димензионих података тако што додељује документе и кластере један у односу на другог тако да се документи са сличним темама појављују један близу другог на патентним мапама. Ово омогућава да се врло лако пронађу најразвијенија подручја у технологији. То такође показује грубе грешке у подацима, тј. патенте који не одговарају пуно на дату тему, али су прикупљени случајно. Ниско-димензиони подаци (структурирани и библиографски подаци) су приказани као бар дијаграми и могу такође помоћи бољем разумевању технолошких области, промена у технолошком развоју, кретању конкуренције и слично [28].

Програмска подршка омогућава визуализацију података са великим бројем променљивих. Подаци се приказују на дводимензионалном графику на такав начин да свака тачка представља један патент, а опције које су доступне кориснику су следеће:

- Приказ ознака патената у виду *PID* ознаке за сваку тачку која представља патент
- Групе патената које су разврстане по одређеном критеријуму (технологија, проналазач и сл.) могу се графички приказати различитим бојама и облицима тачака
- Називи група патената су приказани у легенди, а приказивање свих патената који припадају једној групи може се укључивати и искључивати по потреби
- Дијаграм је могуће увећати
- У сваком тренутку могуће је штампати дијаграм, као и сачувати га у облику *PNG*, *JPEG*, *SVG* или *PDF* датотеке
- Тачке је могуће изабрати (означити) и затим им променити неке од параметара:
 - приказ на графику
 - промена ознаке тачке (ознака тачке је стандардно *PID* патента)
- За означену тачку на графику могуће је приказати следеће библиографске податке:
 - Каталожки број патента (енгл *PID*)
 - Назив патента (енгл *Title*)
 - Проналазач (енгл. *Inventor*)
 - Носилац права (енгл. *Assignee*)
 - Приоритетни датуми (енгл *Priority Date*)
 - датум издавања патента (енгл *Filing Date*)
 - Међународна класификација патената (енгл *IPC*)
 - Кратак текст о патенту - сажетак (енгл. *Abstract*)
- Постоји и могућност да се приступи директно интернет страници патента на сајту организације код које је патент заведен.

Након учитавања података, приказује се дијаграм, на којем су исцртане тачке изнад којих се налазе ознаке патената (Слика 14). Ознаке изнад патената сметају у случајевима када су тачке густо постављене. Ове ознаке се могу прикрити уз помоћ поља за са називом “*Display Point Label?*”. Подацима који се односе на одређену тачку може се приступити постављањем показивача миша изнад тачке, након чега се појављује прозор са информацијама о тој тачки: назив групе, каталошки број, веза ка интернет страници патената на презентацији организације код које је патент заведен и координате тачке.

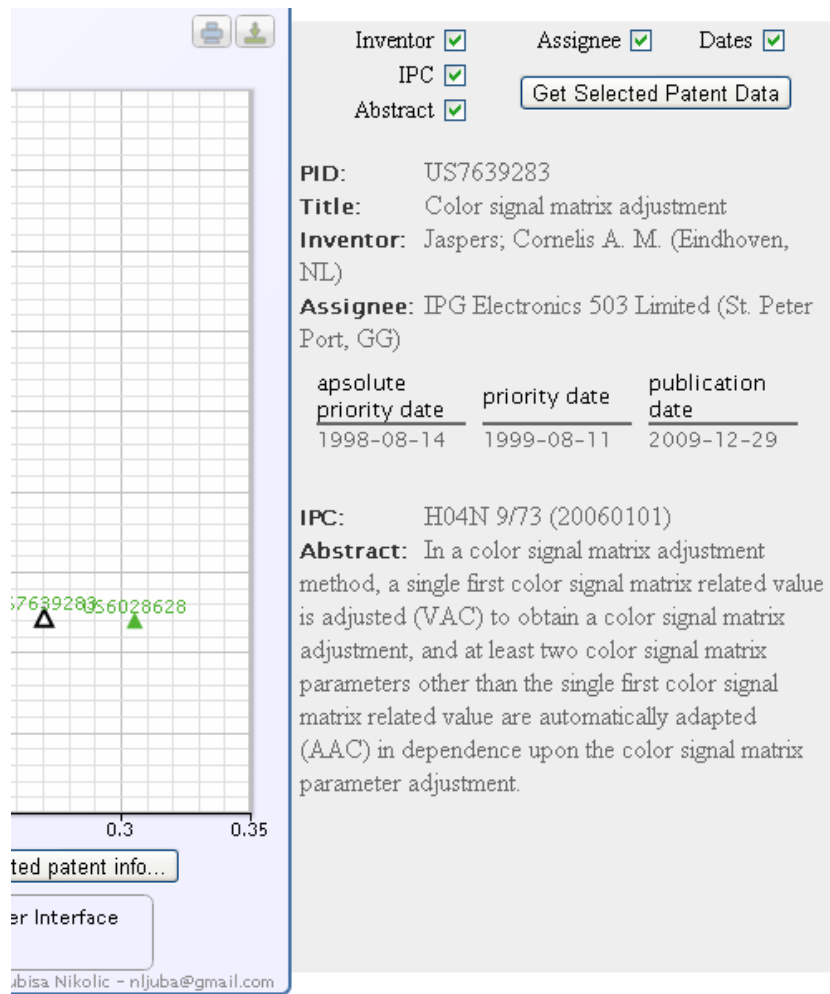
Легенда се налази испод дијаграма и она показује повезаност боје и облика тачке са групом којој она припада. Могуће је прикрити одређену групу патената притиском тастера миша на назив групе у легенди. Након сакривања групе патената, дијаграм се аутоматски увећава како би се преостали патенти приказали што боље на површини графика.

Анализа графика се у великој мери олакшава приказом библиографских података о патенту чија је тачка означена (Табела 6). За означену тачку на графику могуће је приказати следеће библиографске податке:

	Оригинални назив	Објашњење
a)	<i>PID</i>	Каталошки број патената
b)	<i>Title</i>	Наслов патената
c)	<i>Inventor</i>	Проналазач
d)	<i>Assignee</i>	Носилац права
e)	<i>Priority Date</i>	Приоритетни датум
f)	<i>Issue Date</i>	Датум издавања патената
g)	<i>IPC</i>	Међународна класификација патената
h)	<i>Abstract</i>	Сажетак

Табела 6 Библиографски подаци

Корисник може изабрати које податке жели да буду приказани на основу означених поља за потврду (Слика 15), а након тога уз помоћ тастера “*Get Selected Patent Data*” приступа се локалној бази из које се преузимају жељени подаци.



Слика 15 Изглед додатних информација о патенту

Позивање интерфејса за штампање графика врши се тастером са сличицом штампача, који се налази у горњем десном углу графика. Снимање слике графика могуће је обавити у сваком тренутку, а слика се може сачувати у облику *PNG*, *JPEG*, *SVG* или *PDF* датотеке.

6. Резултати и анализа

6.1 Перформансе интернет робота

Испитивање перформанси интернет робота обављено је тестирањем патената који имају различиту количину података, тј. изглед стране са подацима им је потпуно другачији. Тестирани су следећи патенти:

- [7962846](#) - *USPTO* патент са класичним подацима без додатних табела
- [7919816](#) - *USPTO* патент има додатну табелу “*Foreign Application Priority Data*”
- [7962825](#), - *USPTO* има додатне блокове са подацима: “*Related U.S. Patent Documents*” и “*Parent Case Text*”
- [D503691](#) – дизајн патент који има специфичан изглед интернет странице са ограничено много података, има и цитиране патенте
- [D254200](#)

Анализирано се извршавање програма и пролазно време прибављања, парсирања и записивања података у базу, као и статистички подаци о количини података који су записани у базу. Брзина интернет везе која се при тестирању користила је *6Mbps* у долазном саобраћају и *0,36Mbps* у одлазном саобраћају. Брзина интернет везе је важан фактор јер она највише утиче на брзину рада програма.

Табела 7 приказује количину података која треба да се обради по патенту.

PID	Табеле у бази					
	<i>Search Setup</i>	<i>Main Data</i>	<i>Case Data</i>	<i>Citation Data</i>	<i>claims</i>	<i>ipc</i>
7962846	1	17	1	16	516	64
7919816	1	7	1	6	140	26
7962825	1	11	1	10	182	39
D503691	1	18	1	17	317	76
D254200	1	15	1	17	7	16

Табела 7 Статистика прикупљених података за први тестни случај

Табела 8 приказује време потребно за обраду наведених патената.

Ознака	Укупно време	Време по патенту
7962846	23s 827ms	1.40 s
7919816	12s 359ms	1.76 s
7962825	21s 671ms	1.97 s
D503691	23s 235ms	1.29 s
D254200	16s 155ms	1.07 s

Табела 8 Брзина прикупљања једног патента за први тестни случај

Табела 9 представља статистику преузимања једног просечног патент портфолија од нпр. 1820 патената а Табела 10 представља просечно време преузимања. Просечно време је добијено за јединствене патенте, тј патенте који су заиста преузети и уписани у базу, док се дупликати, тј патенти који су у бази већ постојали, не преузимају већ се само повезују са текућом претрагом.

Главних патената	Бр. цитата	Бр. патената који цитирају	Просечна цитираност	Просечно цитираних патената
1820	55754	29121	30.63	16.00

Табела 9 Други тест случај 1820 патената

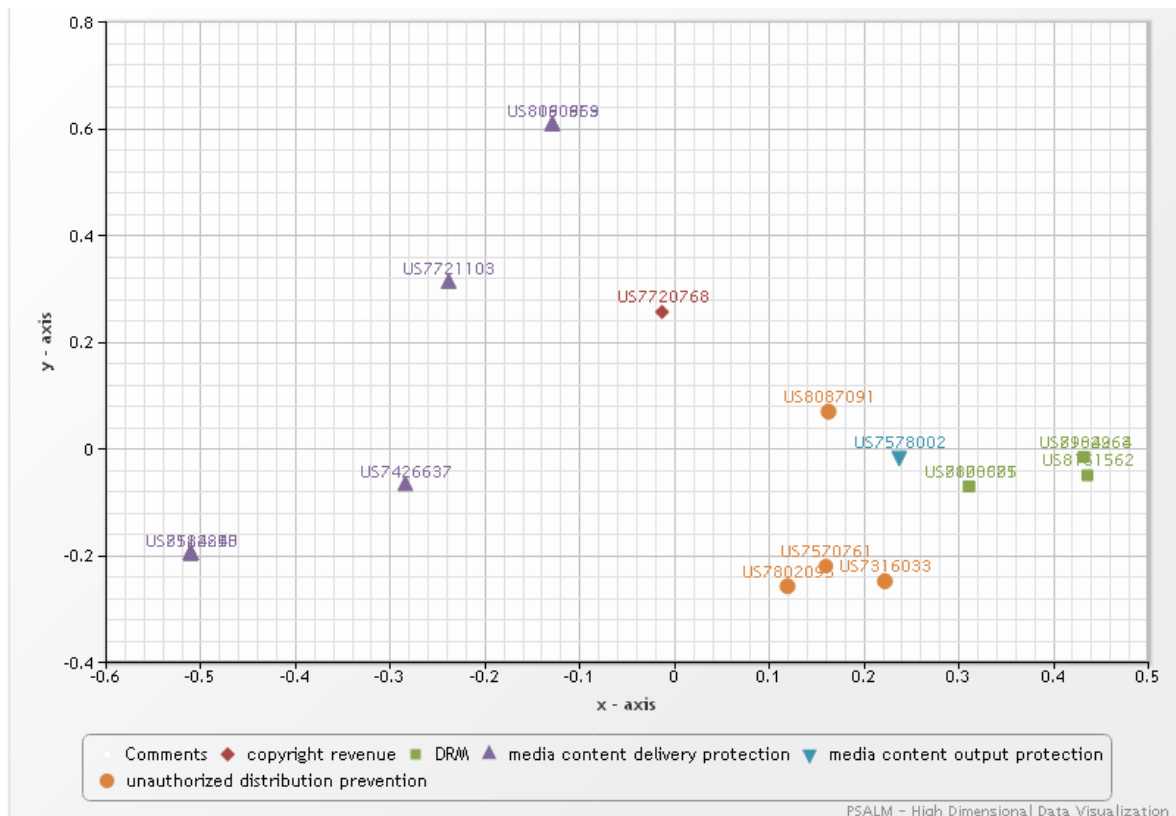
Укупно време рада	Укупан бр. патената	Укупан бр. јединствених патената	Просечно време [s]
17h 15m 42s 567ms	59829	30225	2.06

Табела 10 Тест случај 1820 патената просечно

6.2 Верификација расподеле патената у 2Д простору

Циљ овог теста био је да покаже да је расподела тачака у 2Д простору довољно добра. Верификација се састоји у томе што ће се скуп патената пропустити кроз алгоритам и приказати у 2Д простору а затим ће се ручно груписати на основу мишљења које су претходно дали експерти. На тај начин ће се визуалним путем потврдити добар или лош распоред патената у простору. За ову сврху одабран је мањи скуп од 19 патената које су првобитно експерти прегледали и класификовали. Испоставило се да се могу дефинисати 3 кластера и 2 сингуларитета (патенти који се нису могли класификовати са осталим патентима, већ појединачно представљају кластер). Групе су дефинисане као:

- *Digital Rights Management – DRM (5)*
- *Media Content Delivery Protection (8)*
- *Media Content Output Protection (1)*
- *Copyright Revenue (1)*
- *Unauthorized Distribution Protection (4)*



Слика 16 Верификација расподеле патената у 2Д простору

Слика 16 показује врло добру визуалну расподелу патената по групама, поготово што су поједини патенти који припадају истој групи на дијаграму практично преклопљени, тако да се може закључити да је расподела патената у 2Д простору на задовољавајућем нивоу.

6.3 Тест метода за класификацију података

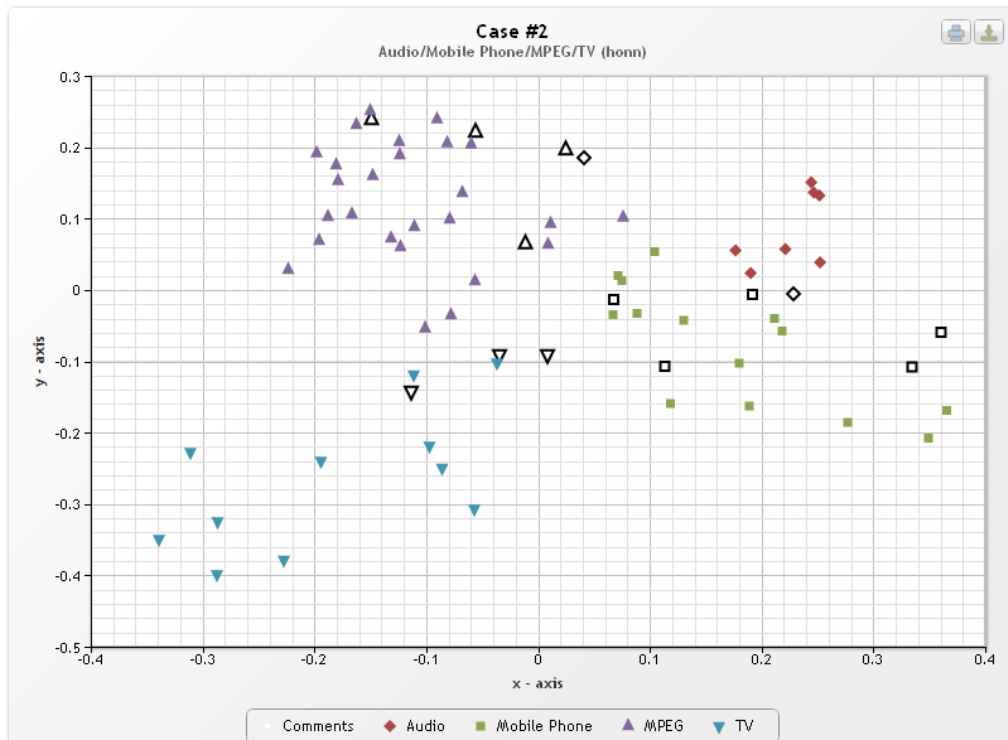
Приликом одабира најпогодније методе за кластеризацију података методе су тестиране на репрезентативном узорку патената. За тестирање метода груписања користиће се 72 америчка патената из свеобухватног скупа тачака. Сви патенти су у власништву исте компаније и покривају области потрошачке електронике. Стручњаци из ове области су груписали ове патенте у следеће четири групе: Аудио (16 патената), Мобилни телефони (15 патената), видео (29 патената) и ТВ (12 патената).

6.3.1 Опште познате методе

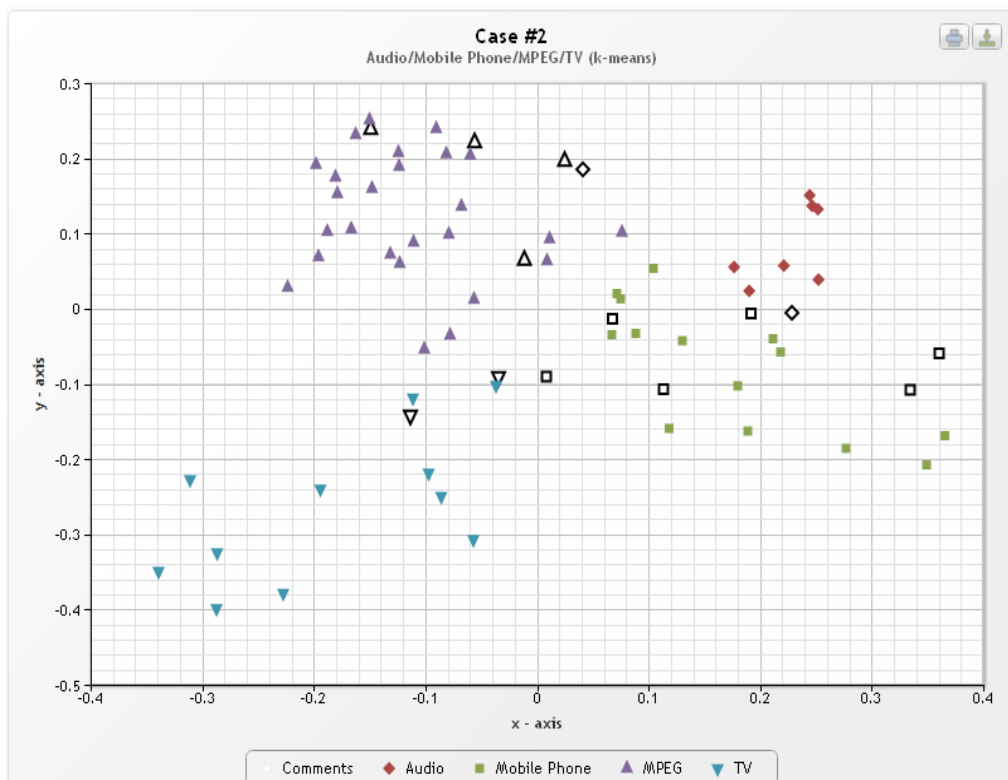
У овом поглављу поредиће се познате и најчешће коришћене методе за класификацију.

6.3.1.1 Визуелно поређење

Да би се визуелно упоредили резултати стручњака и аутоматско груписање добијено алгоритма направљене су патентне карте тзв. пејзажне карте (енгл *landscape map*). Свака група је представљена различитим симболом. „Audio” група је представљена ромбом, „Mobile Phone“ квадратом, „Video“ троуглом и „TV“ обрнутим троуглом. Након тога патентне мапе су упоређене са резултатима стручњака. Слика 17 приказује резултате за „RONN“ алгоритам, а Слика 18 за „*k-means*“ алгоритам. Подебљане тачке представљају неусклађеност између кластера стручњака и кластера одређеног алгоритма груписања.



Слика 17 Резултат експерата и *Ronn* алгоритма



Слика 18 стручна анализа наспрам *k-means* алгоритма

6.3.1.2 Математичко поређење

Тачност класификације четири различите методе је израчуната коришћењем програмског пакета „*Matlab*“ и резултати су приказани у табелама које следе. Прва табела представља тачност класификације за сваку од четири методе по сваком кластеру, док друга табела представља перформансе сваке методе.

Алгоритам	Тачност [%]			
	<i>Audio</i>	<i>Mobile Phone</i>	<i>Video</i>	<i>TV</i>
<i>k-means</i>	47.88	74.80	89.72	90.50
<i>neural-gas</i>	43.87	89.60	90.14	91.67
<i>FCM</i>	43.75	93.33	89.66	91.67
<i>RONN</i>	46.25	66.00	90.41	91.67

Табела 11 Тачност класификације четири алгоритама груписања

Алгоритам	Перформансе		
	<i>Average</i> ($\times 10^{-4}$)	<i>MSE</i>	<i>DB index</i>
<i>k-means</i>	85		0.717
<i>neural-gas</i>	74		0.709
<i>FCM</i>	74		0.719
<i>RONN</i>	75		0.706

Табела 12 Перформансе алгоритама груписања

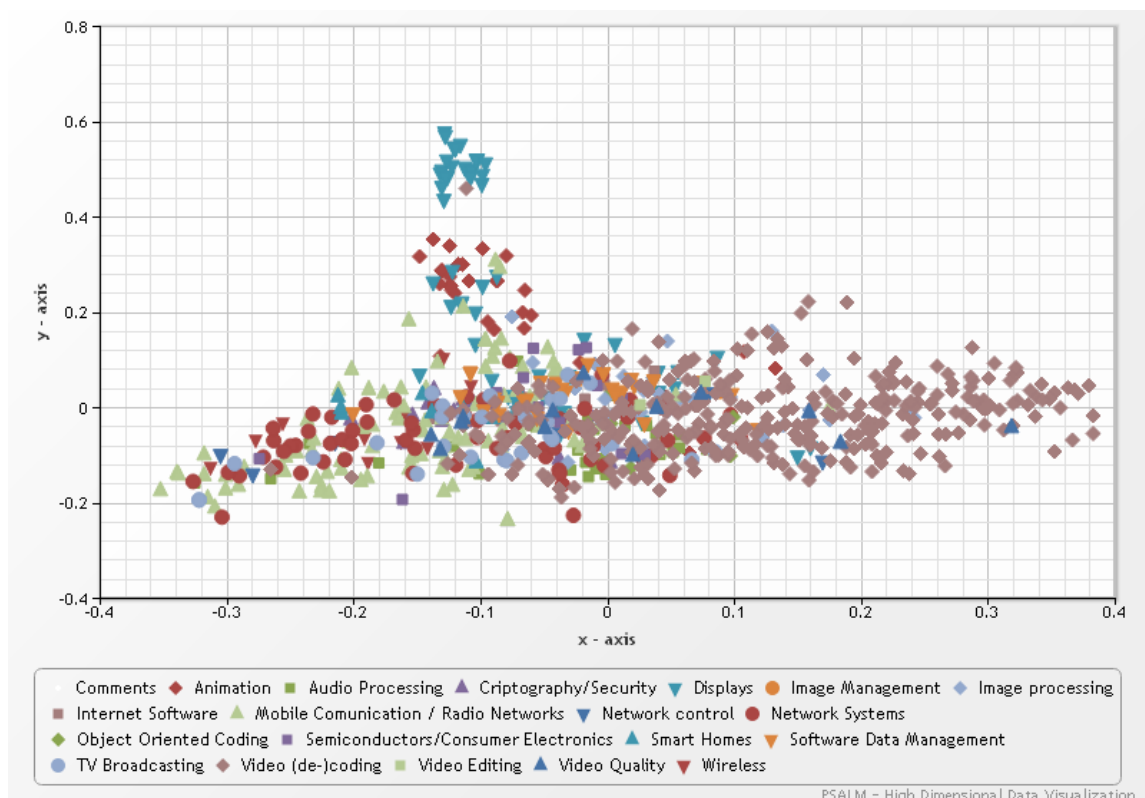
Резултати показују сличне перформансе за све 4 тестиране технике. Аудио група је груписана са мање од 50% прецизности што је веома лоша тачност. Главни проблем не лежи у груписању већ у фази „копања по подацима“. На основу добијених резултата перформанси може се закључити да су све методе врло сличне по том критеријуму.

6.3.2 Властите методе

Како би се проверила тачност програмске подршке иста је тестирана на скупу патената (817) који су претходно прегледали, класификовали и рангирани експерти. Патенти углавном припадају области потрошачке електронике и у власништву су два правна лица. Оцена степена припадности групи процењиваће се по једначини (2), а за N ће се узети вредност 2, што представља релативно строг услов. Визуална оцена постизаће се на основу могућности програмске подршке да када се курсор миша постави на једну од категорија у легенди приказа, на дијаграму се издвајају само патенти који припадају датој групи, док сви остали патенти постају осенчени. На тај начин се омогућава бољи увид у распоред појединих група у односу на остатак дво-димензионог простора патената.

6.3.2.1 Тест методе „груписање по технологијама“

Применом методе груписања улазни подаци су груписани у 19 кластера и то по припадности одређеној технологији (Слика 19).



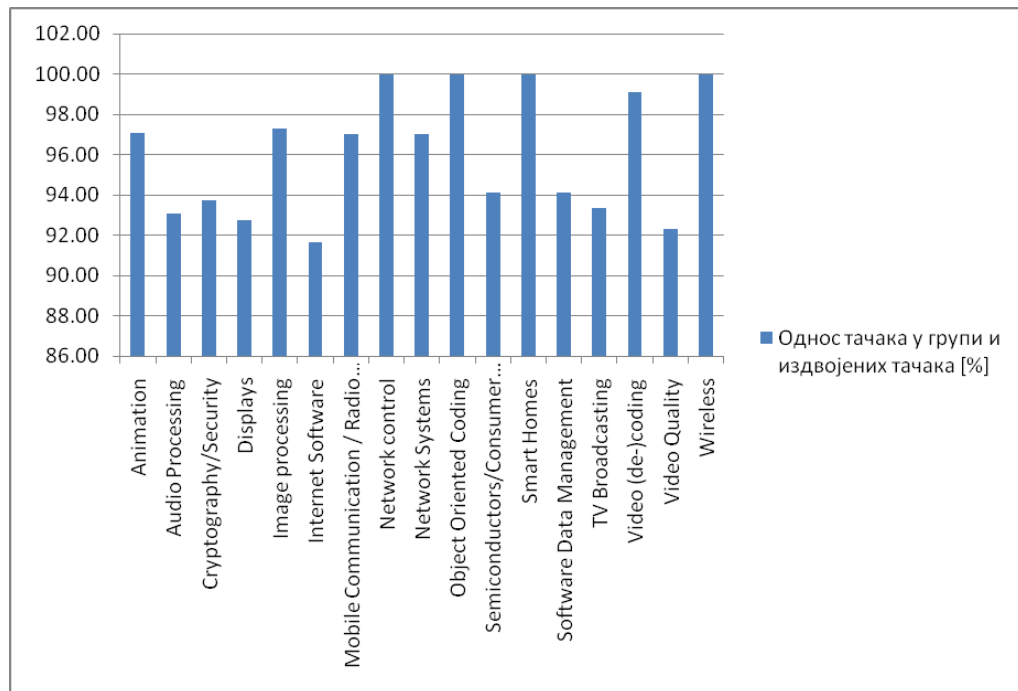
Слика 19 Резултат примене методе „груписање по технологијама“

6.3.2.1.1 Математичко поређење

Табела 13 приказује резултате провере тачности методе коришћењем једначине (2). У последњој колони дат је однос укупног броја тачака и тачака издвојених у односу на центар групе у процентима. Слика 20 приказује добијене резултате графичким путем.

Група	Број патената	Тачност	
		Број издвојених тачака у односу на центар групе	Однос укупног броја тачака и издвојених тачака [%]
<i>Animation</i>	34	1	97.06
<i>Audio Processing</i>	29	2	93.10
<i>Cryptography/Security</i>	16	1	93.75
<i>Displays</i>	55	4	92.73
<i>Image processing</i>	37	1	97.30
<i>Internet Software</i>	12	1	91.67
<i>Mobile Communication / Radio</i>	101	3	97.03
<i>Network control</i>	8	0	100.00
<i>Network Systems</i>	67	2	97.01
<i>Object Oriented Coding</i>	4	0	100.00
<i>Semiconductors/Consumer Electronics</i>	17	1	94.12
<i>Smart Homes</i>	10	0	100.00
<i>Software Data Management</i>	34	2	94.12
<i>TV Broadcasting</i>	30	2	93.33
<i>Video (de-)coding</i>	336	3	99.11
<i>Video Quality</i>	13	1	92.31
<i>Wireless</i>	11	0	100.00

Табела 13 Провера тачности кластер методе

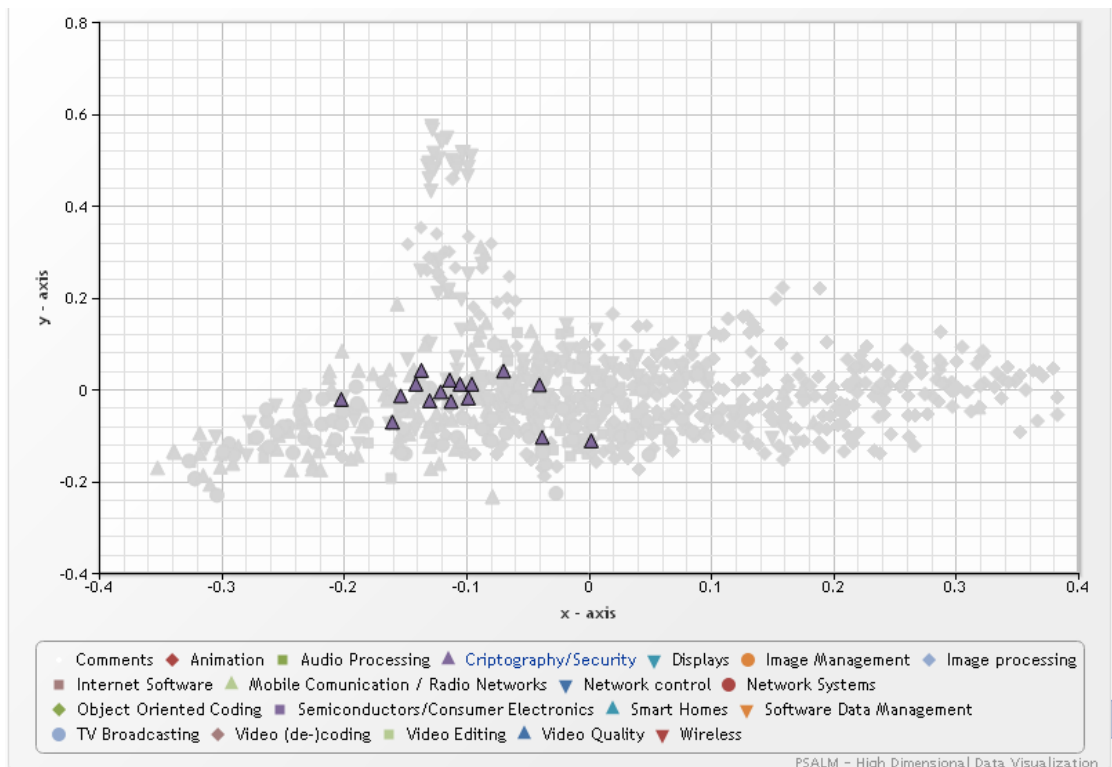


Слика 20 Приказ односа тачака у групи и тачака издвојених из групе

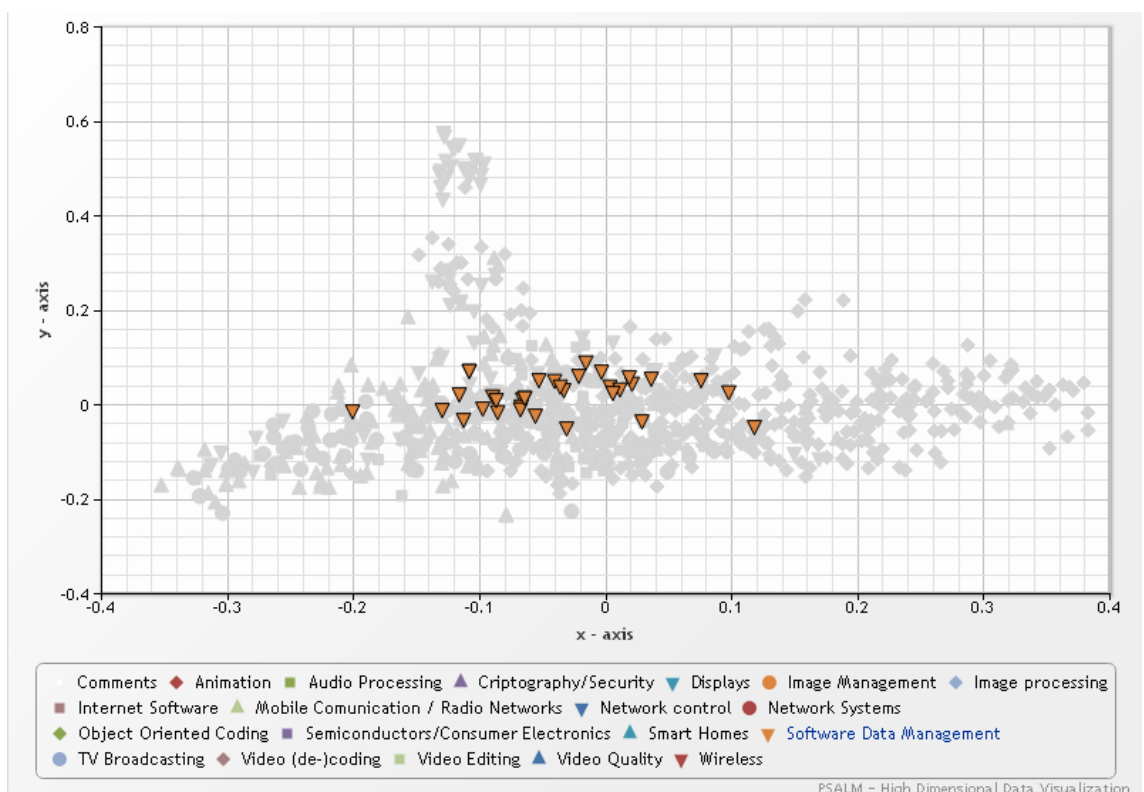
Из приложеног се види да се тачност креће између 90% и 100%, што је веома висока тачност. На основу добијених резултата може се закључити да је метода дала врло добар резултат груписања.

6.3.2.1.2 Визуелно поређење

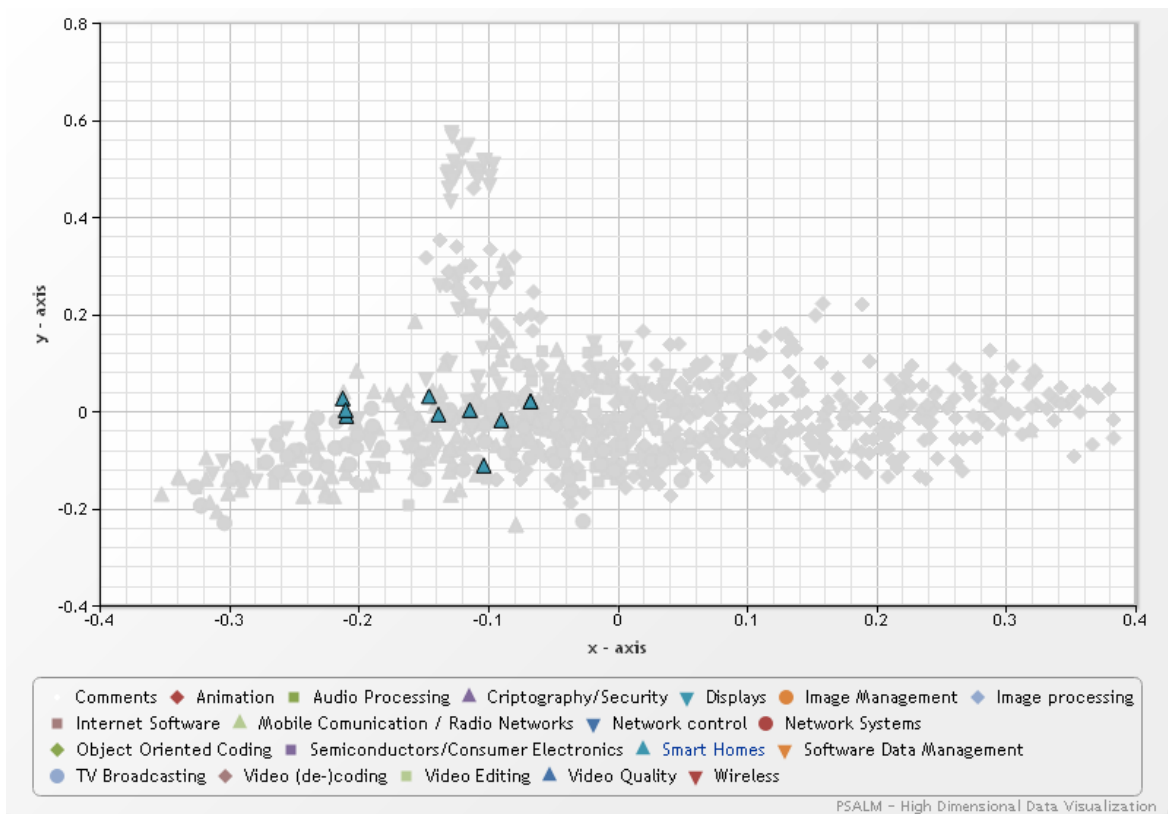
Ако би се неке од група представиле визуално то би изгледало као на сликама које следе. Слика 21 приказује случај преласка курсором миша преко групе означене троуглом („*Cryptography/Security*“). Слика 22 приказује издвајање групе означене обрнутим троуглом („*Software Data Management*“) Слика 23 издвајање групе означене троуглом („*Smart Homes*“) и Слика 24 издвајање групе означене обрнутим троуглом („*Wireless*“).



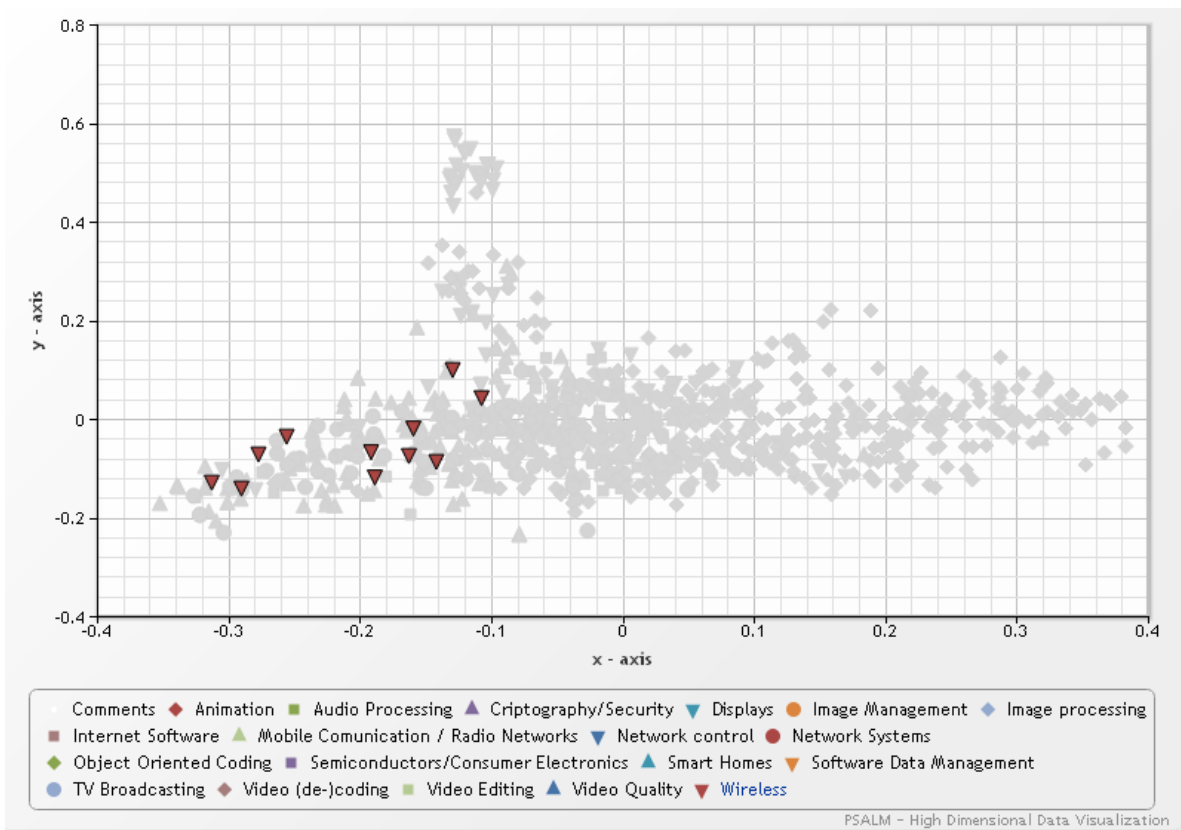
Слика 21 Позиционирање групе „*Cryptography/Security*“



Слика 22 Позиционирање групе „*Software Data Management*“



Слика 23 Позиционирање групе „Smart Homes”



Слика 24 Позиционирање групе „Wireless”

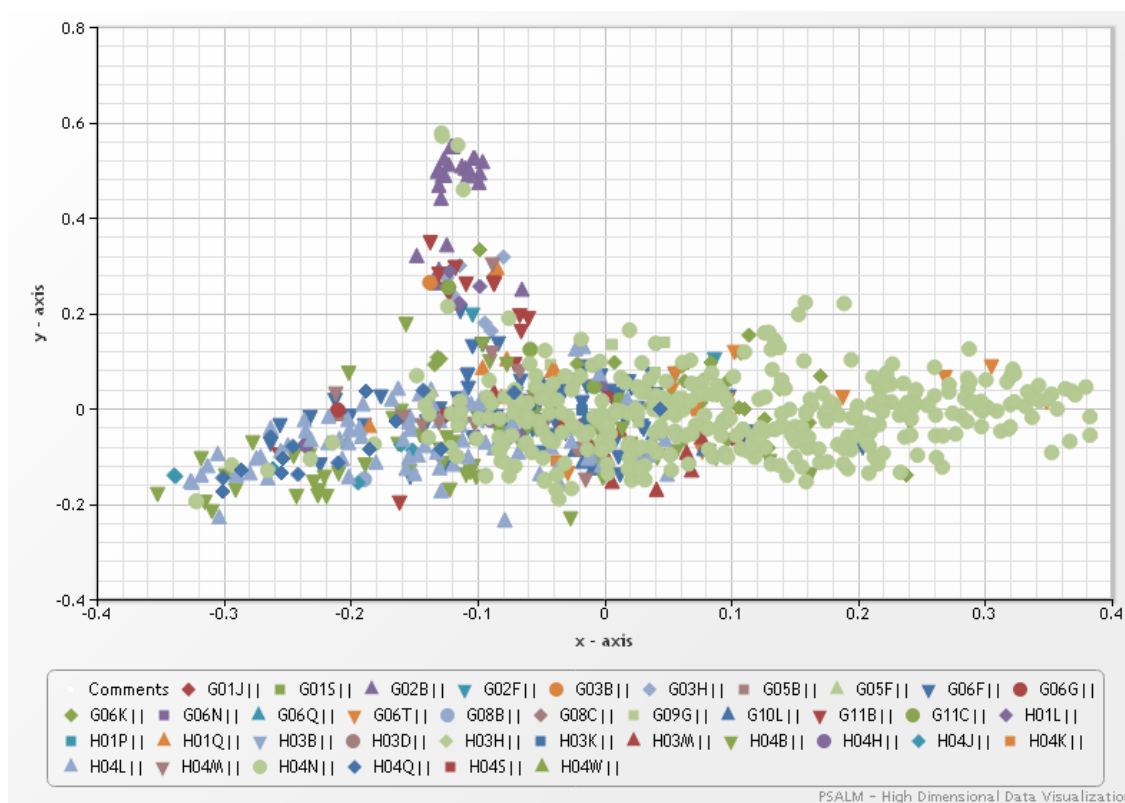
На основу претходних слика може се рећи да визуелни приказ такође потврђује резултате математичког теста - да су групе прилично добро груписане и да нема исувише расутих тачака.

Ови резултати додатно потврђују и прилично добру расподелу патената у 2Д простору, тако да су овом методом успешно повезани резултати груписања (развијене методе формирања група на бази анализе *IPC* кодова и предефинисаних технологија) и расподеле патената у 2Д простору (на основу текста патента и издвојених кључних речи).

6.3.2.2 Тест методе „груписање по кодovima“

Како би се метода „груписање по кодovima“ могла квантитативно и квалитативно упоредити са претходном методом („груписање по технологијама“), за улазне податке узет је исти скуп патената (817). Кориштена је и иста метрика (једначина (2), где је $N=2$) за квантитативну оцену, а такође и визуелна представа добијених скупова.

Применом методе „груписање по кодovima“ улазни скуп података подељен је у дупло више група у односу на методу „груписање по технологијама“ - 38. Називи група су заправо прва четири карактера *IPC* кода (Слика 25).



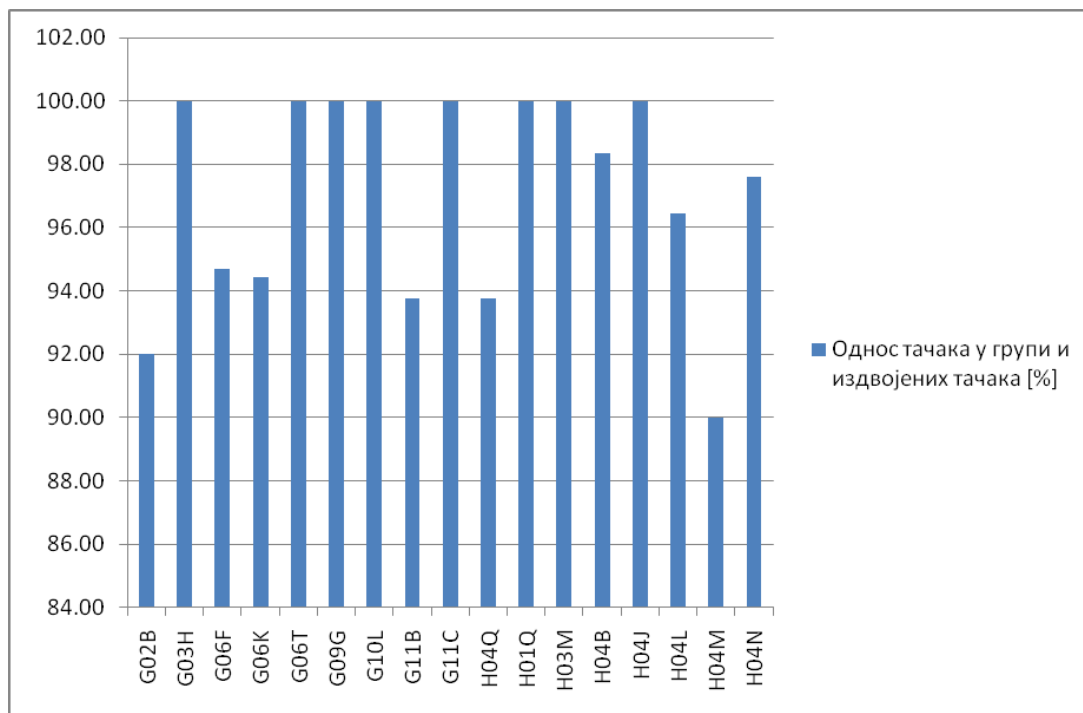
Слика 25 Резултат примене методе „груписање по кодovima“

6.3.2.2.1 Математичко поређење

Табела 14 приказује резултате провере тачности методе „груписање по кодovima“, а Слика 26 приказује добијене вредности графичким путем.

Група	Број патената	Тачност	
		Број издвојених тачака у односу на центар групе	Однос издвојених тачака у односу на центар групе [%]
G02B	25	2	92.00
G03H	6	0	100.00
G06F	94	5	94.68
G06K	36	2	94.44
G06T	16	0	100.00
G09G	7	0	100.00
G10L	7	0	100.00
G11B	16	1	93.75
G11C	5	0	100.00
H04Q	16	1	93.75
H01Q	7	0	100.00
H03M	13	0	100.00
H04B	60	1	98.33
H04J	7	0	100.00
H04L	84	3	96.43
H04M	10	1	90.00
H04N	378	9	97.62

Табела 14 Провера тачности кластер методе

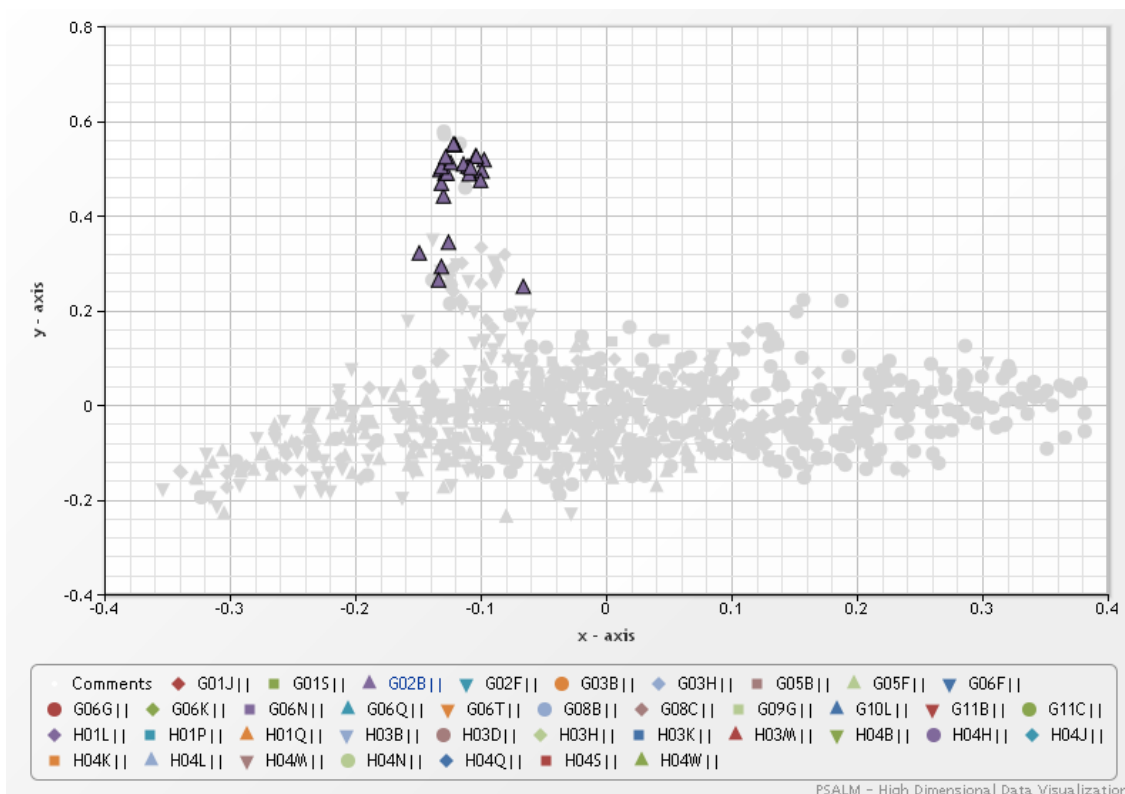


Слика 26 Приказ односа тачака у групи и тачака издвојених из групе

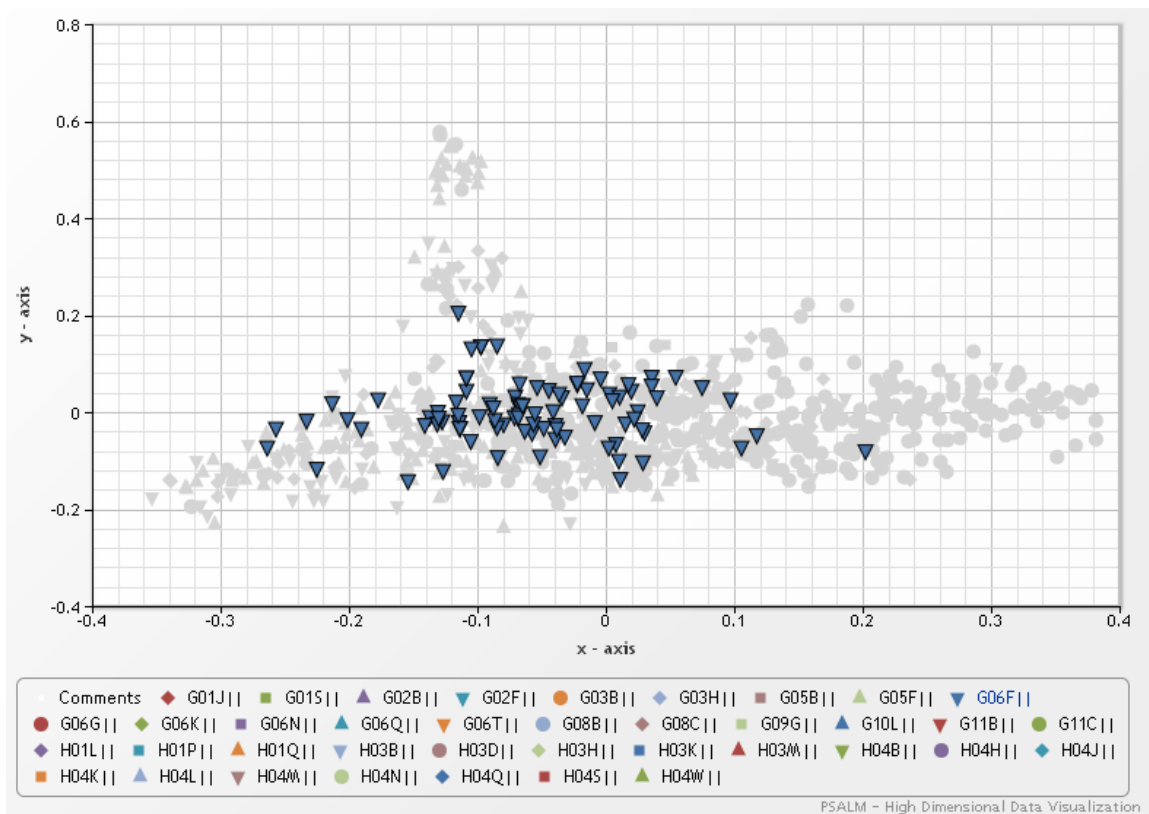
6.3.2.2.2 Визуелно поређење

Слика 27 случај преласка курсором миша преко групе означене троуглом „G02B“ која по класификационом коду представља све уређаје, системе и елементе везане за

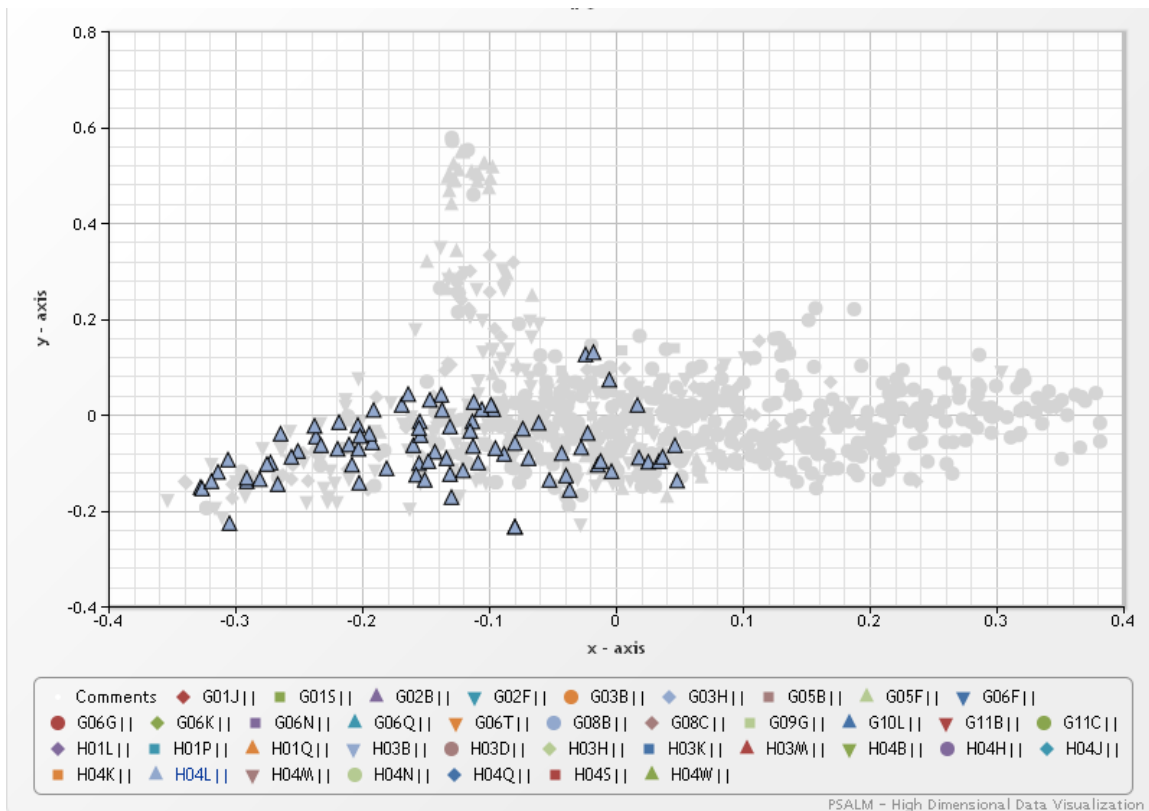
оптику, где се под оптиком подразумева не само видљива, већ и инфрацрвена и ултравиолетна светлост. Слика 28 приказује издвајање групе означене обрнутим троуглом „G06F“, која представља дигиталне процесоре за обраду података. Слика 29 приказује издвајање групе означене троуглом „H04L“ која представља пренос дигиталних информација. Слика 30 приказује издвајање групе означене кругом „H04N“ која по класификацији представља телевизију.



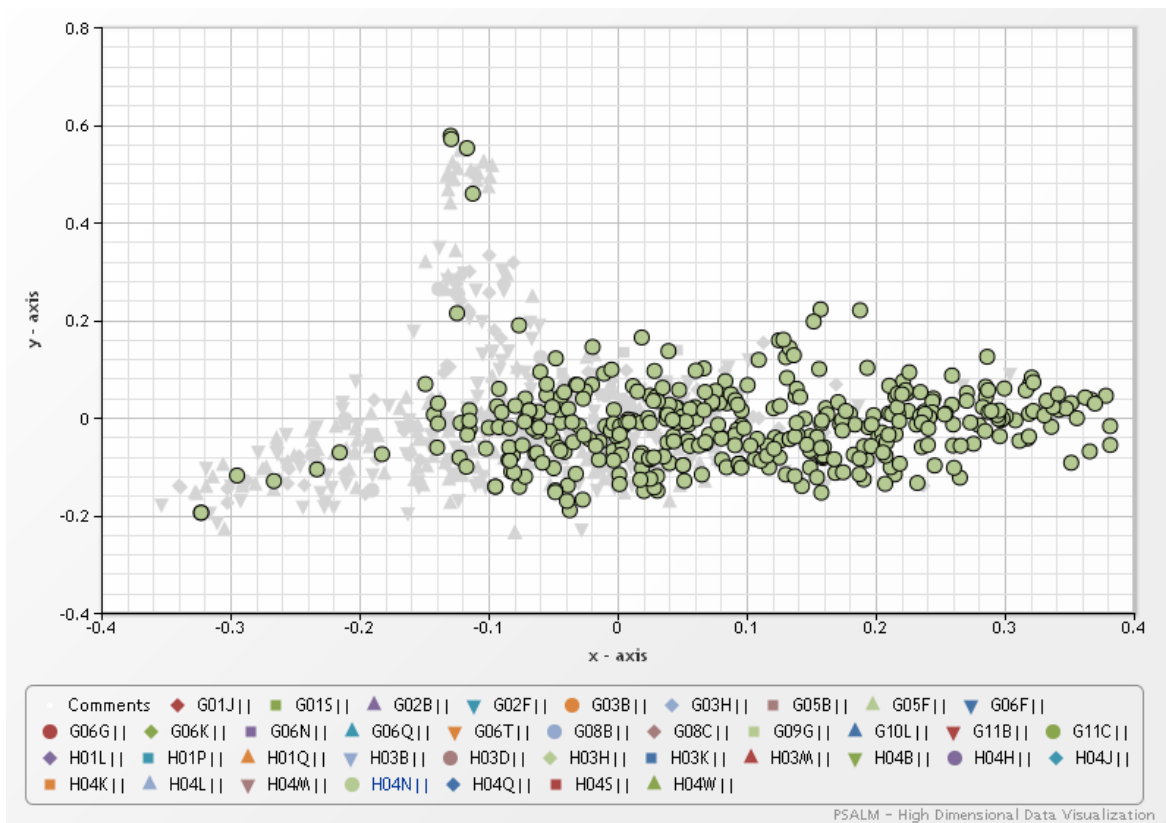
Слика 27 „G02B“ – уређаји, системи и елементи везани за оптику



Слика 28 „G06F“ - дигитални процесори за обраду података



Слика 29 „H04L“ – пренос дигиталних информација



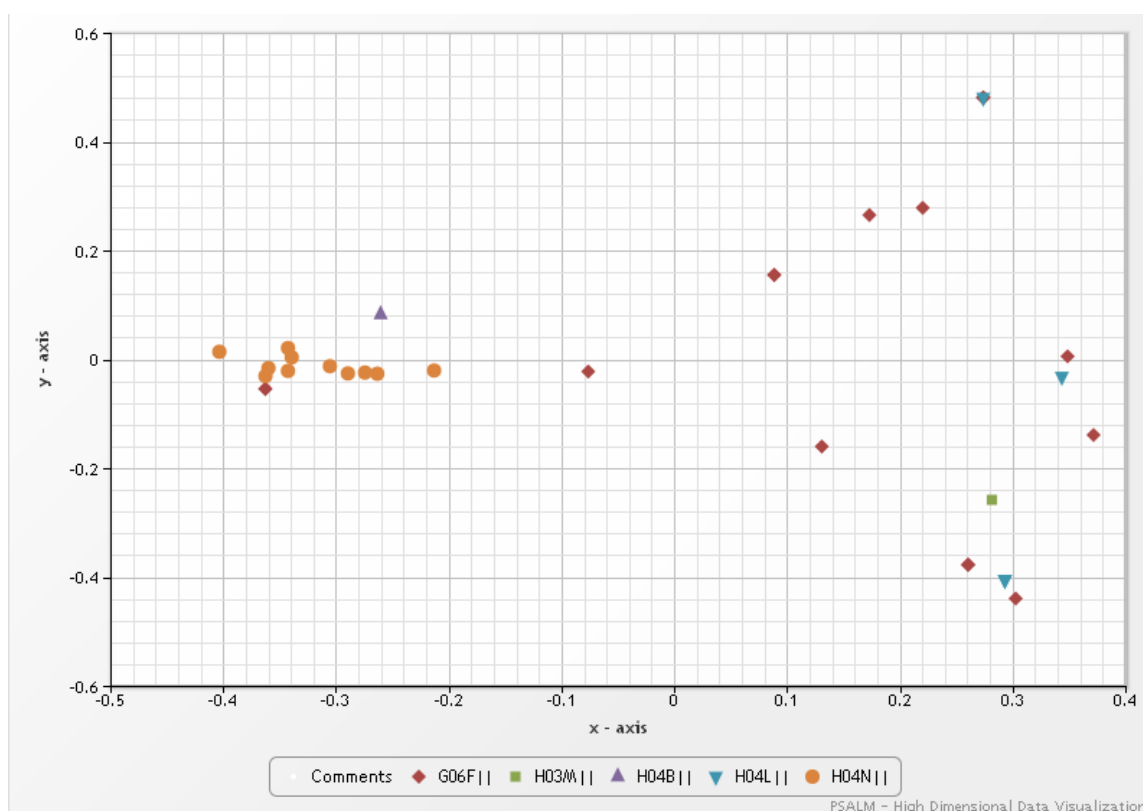
Слика 30 „H04N“ - телевизија

Визуелни приказ потврђује резултате математичког теста да су групе прилично добро груписане. Може се такође закључити да је најзаступљенија група „H04N“ (заступљена у око 40% патената) и покрива комплетан простор технологија Слика 30. Како ова класа представља телевизију, може се рећи на основу груписања по технологијама да заиста велики број патената јесте применљив у овој области, гледајући у ширем смислу.

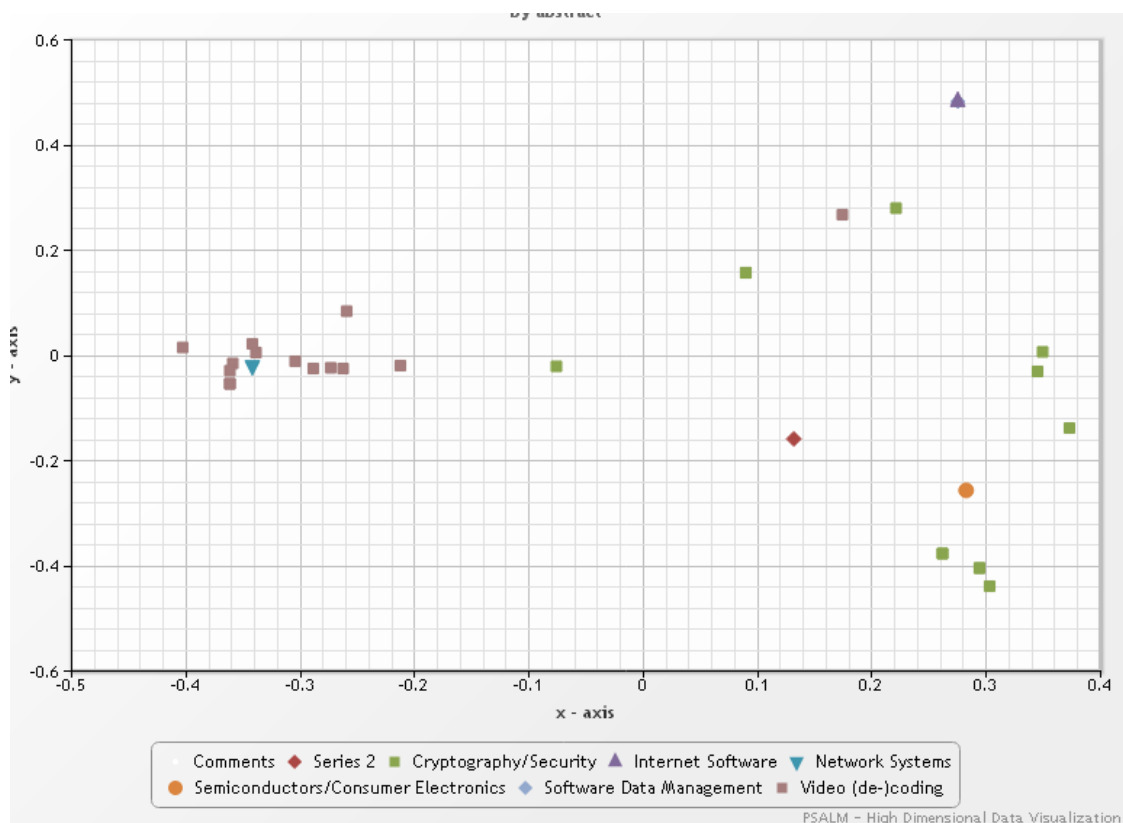
6.3.2.3 Закључак

На основу упоредног теста метода „груписање по технологијама“ и „груписање по кодовима“ могло би се закључити да обе методе дају једнако добре резултате са становишта квантитативног груписања патената у групе које немају пуно издвојених тачака у односу на број тачака у групи. Међутим методом „груписање по кодовима“ добијено је дупло више група у односу на методу „груписање по технологијама“, што и не би био проблем да су групе приближно једнаке величине, али резултати показују велику не сразмеру међу величинама група. Постоје доста велике групе са великим бројем патената (као што је „H04N“ са 378 патената) и прилично велик број малих група (са по једним, два, три или четири патента)- чак 21. Са друге стране „груписање по технологијама“ је дало мали број група са малим бројем патената (свега 3 групе), остале групе имају релативно добру расподелу патената по групама (највећа група има 101

патент „*Mobile Communication / Radio*“). Ако се узме у обзир и ова анализа могло би се закључити да је, укупно гледано, метода „груписање по технологијама“ дала боље резултате. Недостатак ове методе у односу на методу „груписање по кодовима“ је у томе што захтева обучавање, тако да њена тачност у крајњој линији зависи од величине и избора обучавајућег скупа. У општем случају могло би се десити да ова метода остави одређен број патената без одређивања групе, тј. не груписане. Као потврда ове констатације може се посматрати следећи случај. Посматран је нов скуп података од 19 патената који су углавном из области дистрибуције мултимедијалних садржаја којима су придодати још патенти везани за декодирање и кодирање видеа – њих 15. Слика 31 показује резултате груписања методом „груписање по кодовима“, а Слика 32 резултате добијене методом „груписање по технологијама.“



Слика 31 Резултат груписања на основу првог дела кода



Слика 32 Груписање по технологијама

Слика 32 показује да један патент није могао бити класификован, јер за његове класификационе кодове не постоји предефинисана технологија (патент означен ромбом и називом групе “Series 2“).

6.4 Употребна вредност алата у реалним случајевима

Употребна вредност алата биће приказана на конкретним примерима и реалним портфолијима компанија. Кроз ове примере биће приказане разне могућности програмске подршке и како она може допринети бољем доношењу одлука.

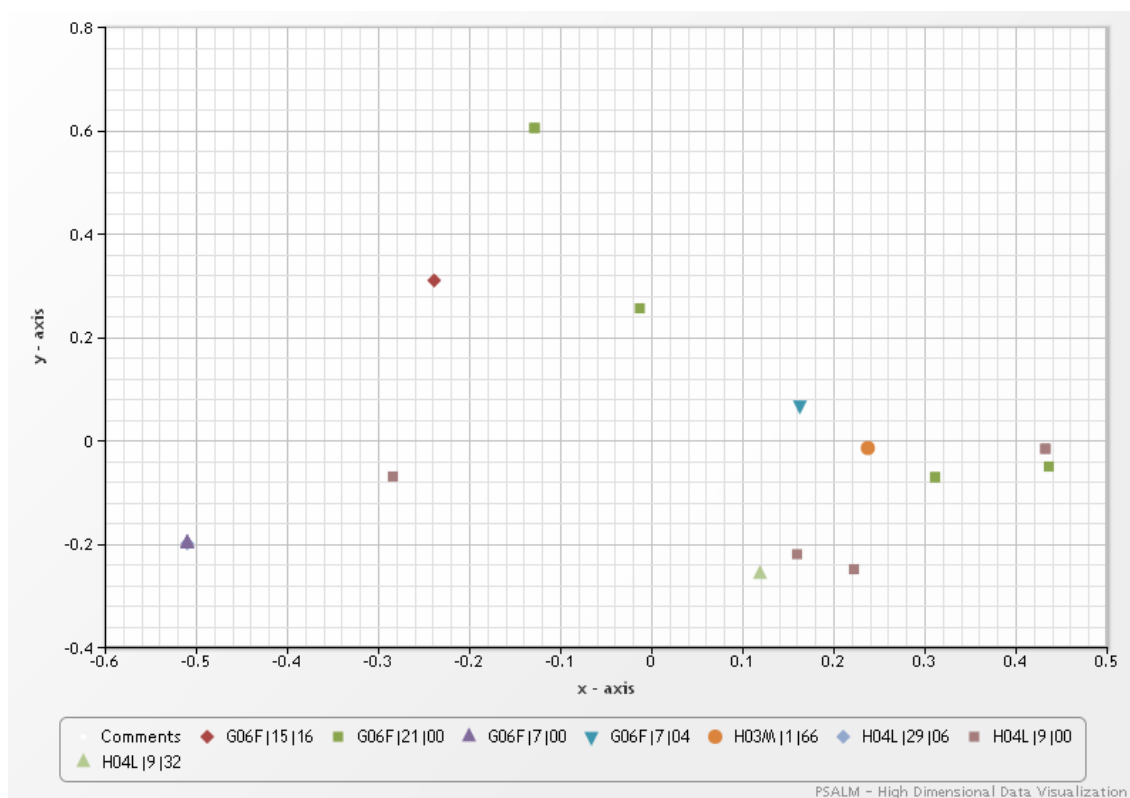
6.4.1 Оцена јачине портфолија компаније

Анализа и оцена јачине патентног портфолија неке компаније показана је на примеру где скуп података чине 19 патената који су углавном из области дистрибуције мултимедијалних садржаја. Наиме, задатак је да се процени колика је снага овог портфолија у односу на портфолио компаније „Microsoft“. Анализа је текла следећим корацима:

- покренут је интернет робот да претражи и ускладишти 19 посматраних патената.
- анализирани су оригинални патенти – 19 помоћу методе „груписање по кодовима“

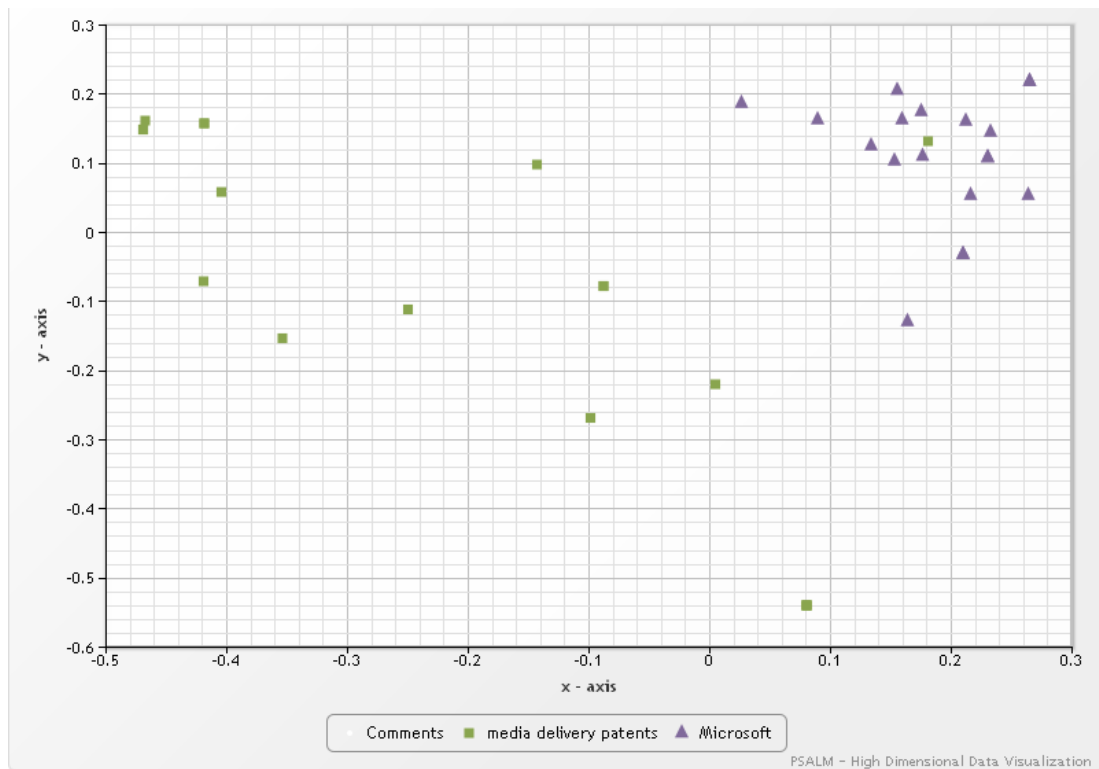
- пуштен је интернет робот да претражи патенте компаније „Microsoft“ по два најзаступљенија кода и то једном користећи логичко „И“ међу кодовима, а други пут користећи логичко „ИЛИ“.
- спојена су два и два случаја претраге
- анализирани су пронађени патенти компаније „Microsoft“ наспрам оригиналних патената у оба случаја.

Након претраге и складиштења 19 оригиналних патената и покретања методе „груписање по кодовима“ издвојена су два водећа класификациона кода – „G06F21/00“ и „H0419/00“. Слика 33 показује да су „G06F21/00“ са 6 и „H0419/00“ са 5 патената, најзначајнији (најзаступљенији) класификациони кодови у овој категорији (групи) патената.



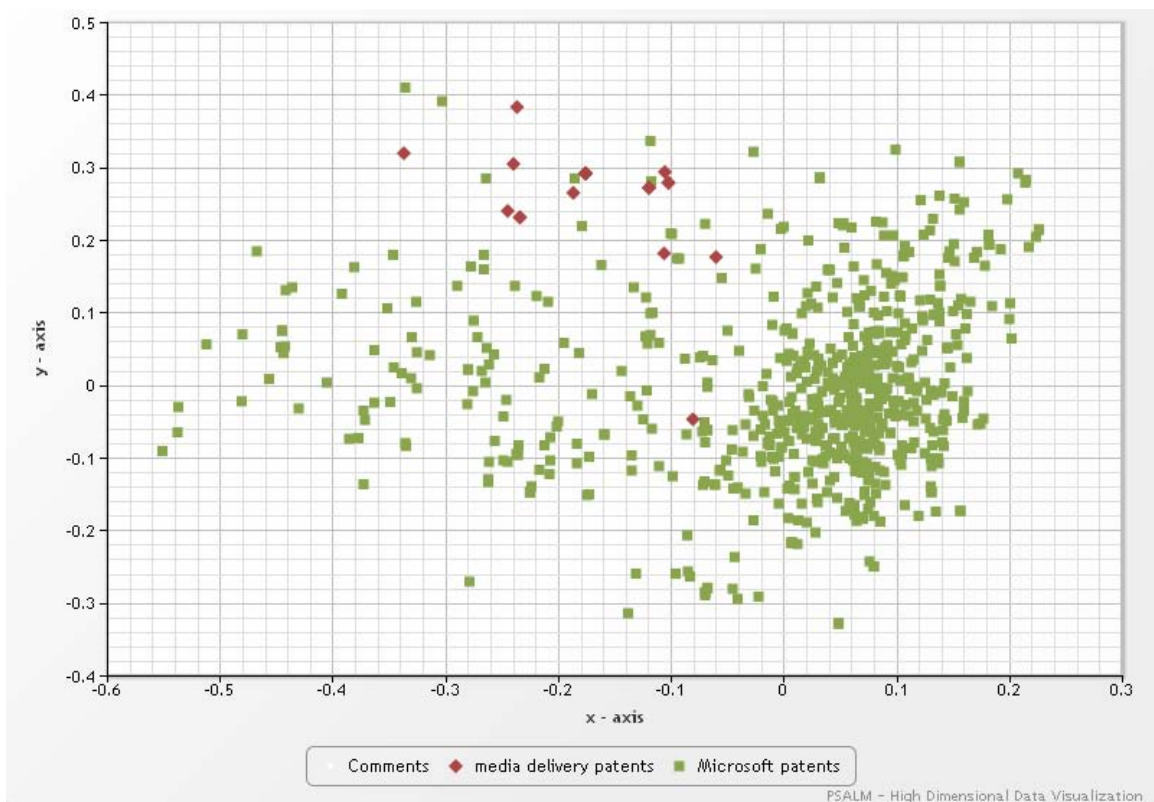
Слика 33 19 патената груписаних по кодовима до трећег нивоа

Стога је покренут интернет робот са комбинацијом „(ICL/G06F21/00 OR ICL/H0419/00) AND AN/Microsoft“ и са комбинацијом „ICL/G06F21/00 AND ICL/H0419/00 AND AN/Microsoft“. На први захтев интернет робот је ускладиштио 726 патената који одговарају задатом критеријуму, а на други захтев 19. Слика 34 показује однос 19 оригиналних патената наспрам 19 пронађених патената компаније „Microsoft“. Са дијаграма се види да су патенти прилично добро подељени у 2Д простору сем једног патената који се нашао у групи патената компаније „Microsoft“.



Слика 34 19 патената у простору са 19 патената компаније „Microsoft“

Слика 35 такође потврђује да су патенти компаније „Microsoft“ концентрисани у једном делу 2Д простора где се не налазе оригинални патенти, а у простору оригиналних патената су мање заступљени.



Слика 35 19 патената у простору са 726 патената компаније „Microsoft“

Из приложеног би се могло закључити да иако компанија „*Microsoft*“ има велик број патената из исте датих класа и да је њена снага у области прилично велика (на основу броја патената), ипак она нема блиско везаних патената оригиналним патентима (према расподели у 2Д простору), тј. да нема патената који су уско повезани са оригиналним. Патент који је у претходном случају био најближи патентима компаније „*Microsoft*“ (једини патент означен квадратом у групи са троугловима) и у овом случају је најближи групи патената компаније „*Microsoft*“ са најгушћом концентрацијом (црвени ромб међу густо распоређеним квадратима).

На основу добијених резултата могло би се закључити да иако компанија „*Microsoft*“ има доста велик број патената из исте технолошке области у којој се налази и већина оригиналних патената, они ипак нису преклопљени у 2Д простору што значи да технолошки нису довољно блиски и да компанија „*Microsoft*“ не би могла квалитетно да се супротстави неком од оригиналних патената ако би се нашли у спору.

6.4.2 Анализа судских спорова

Као један од посебно интересантних резултата овога рада је повезивање података о патентима са подацима о стању и кретањима на тржишту код конкурентних компанија. У том смислу истраживан и анализиран је случај актуелних парница против Андроид-а (оперативног система развијеног првенствено за тзв. паметне телефоне).

Како су уређаји засновани на Андроид ОС-у почели узимати већи удео у расподели добити на светком, а и тржишту у Сједињеним Америчким Државама, тако су почеле и пристизати тужбе у америчким федералним судовима против произвођача који су на том тржишту продавали паметне мобилне телефоне засноване на Андроид ОС-у и то, наравно, првенствено од два главна конкурента компанија: „*Apple*“ и „*Microsoft*“. Отпочет је практично прави рат усмерен против Андроид оперативног система.

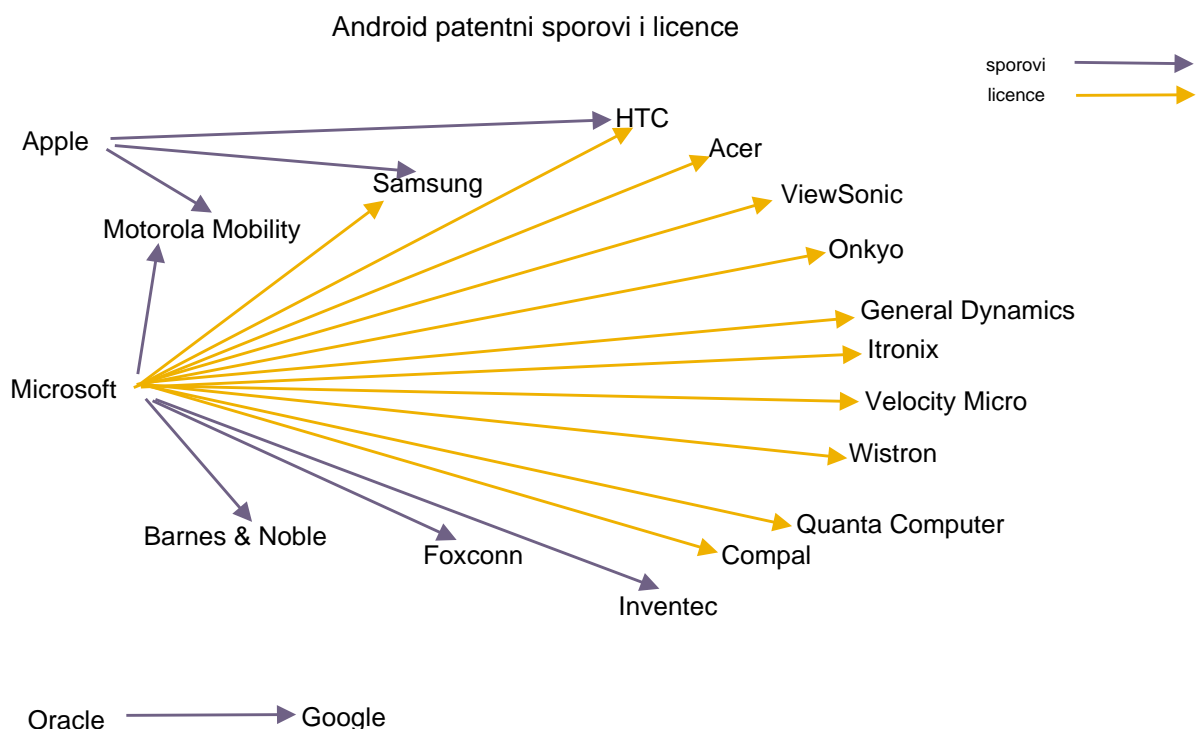
„*Apple*“ се определио за стратегију којом жели спречити друге произвођаче да користе њихове изуме. „*Apple*“ не жели да убира приход на име лиценци које би им компаније плаћале да за узврат користе њихова решења у својим производима, већ жели да издејствује забране продаје оних производа који нарушавају њихове патенте. Вођена тиме компанија „*Apple*“ је покренула неколико тужби против компанија који представљају главне конкуренте као што су „*HTC*“, „*Samsung*“ и „*Motorola Mobility*“.

„*Microsoft*“ се определио за стратегију убирања прихода на основу лиценци које би им компаније плаћале. Тренутно, велики број компанија плаћа лиценце за патенте које нарушава, а у власништву су „*Microsoft*“-а. На тај начин „*Microsoft*“ већ убира велики део прихода на основу лиценци. Та компанија тврди да више од 70% свих телефона

заснованих на Андроид оперативном систему који се продају у Сједињеним Америчким Државама плаћа лиценце за њихов патентни портфолио.

За разлику и од компанија “Apple” и “Microsoft”, компанија “Oracle” опет има трећу стратегију у обрачуна са Андроид оперативним системом. Наиме, “Oracle” је једна од ретких компанија које се одлучила да тужи директно “Google” за нарушавање неких од патената који су у власништву компаније “Oracle”. “Oracle” тужи “Google” због коришћења Јаве (енгл *Java*) у Андроид оперативном систему од 2011 године. Иницијални захтеви компаније “Oracle” се односе на плаћање одштете од око 6 милијарди америчких долара на име компензације за учињену штету и такси за ауторска права. Касније су ови иницијални захтеви спуштени и највероватније ће се задржати на нивоу можда и мањем од 100 милиона америчких долара. Можда још већи проблем за компанију “Google” је то што компанија “Oracle” може да затражи забрану дистрибуције Андроид оперативног система.

Слика 36 приказује стање у споровима који се воде закључно са првим половином 2012 године и стање у плаћању накнаде на основу лиценцих права а све усмерено против компанија које користе Андроид у својим производима. Види се да је број компанија значајан, поготово за случај плаћања накнада.



Слика 36 Спорови и лиценце у вези са Андроид оперативним системом

Ако бисмо се запитали у ком односу се тренутно налазе судски спорови и уговори о лиценцама међу компанијама, тада бисмо добили следећи однос: судски спорови износе једва око 1%, а лиценци уговори око 99% укупног трговања на тржишту интелектуалне

својине. Недавно је “Microsoft” забележио значајну победу у тзв. рату на пољу мобилних телефона, тако што је успео да убеди компанију „Samsung“ да плати ауторска права на сваки уређај који Јужно-Корејска компанија прода, а који је заснован на Андроид оперативном систему. Спремност компаније „Samsung“ да плати ауторска права компанији “Microsoft” која тврди да Андроид нарушава одређена права из њихове интелектуалне својине, такође намеће питање колико су ефикасни напори компаније “Google” да обезбеди правну заштиту произвођачима хардвера који користе Андроид оперативни систем? Компанија “Microsoft” је такође осигурала исплату на име ауторског права од тајванске компаније „HTC“ и то, по незваничним подацима, у износу од 5\$ по једном продатом уређају. Како је Андроид постао водећи оперативни систем за паметне телефоне, предвиђа се да ће до 2015 године годишња продаја уређаја заснованих на овом оперативном систему достићи 430 милиона комада. То би потенцијално донело заиста велику финансијску корист компанији “Microsoft”. Додатно, компанија “Microsoft” је тражила да се ауторска права плаћају и за уређаје попут тзв. таблет-а и електронских читача, а могла би се у будућности пренети и на друге електронске уређаје [29].

У том рату који се води на пољу интелектуалне својине “Google” је недавно изгубио још једну битку. Од својих директних конкурената, компанија: “Apple” и “Microsoft”. У конзорцијуму у којем је учествовало неколико компанија, „Google” је изгубио битку у куповини преко 6000 патената који су припадали компанији „Nortel”. „Google” је понудио 4,5 милијарди америчких долара, што није било довољно. Из тог разлога „Google” је тренутно у поступку преузимања компаније „Motorola Mobility” уз коју би добио и њену интелектуалну својину како у домену мобилне телефоније тако и у домену дигиталне телевизије. „Google” је понудио око 12,5 милијарди америчких долара у августу 2011. године за „Motorola Mobility” уз коју би добио око 17000 већ одобрених патената и још око 7500 патентних пријава. Ова куповина је вероватно била мотивисана тиме што је „Motorola“ покренула парницу против једног од највећих конкурената компаније „Google” - компаније „Apple” са извесним изгледима да ту парницу и добије. Међутим, у међувремену, „Motorola“ је изгубила тај спор против компаније „Apple“, па се поставља суштинско питање: да ли је компанија „Google“ урадила прави потез када је потрошила 12,5 милијарди америчких долара да би присвојила патенте компаније „Motorola“ [30]. Овај случај показује да није довољно купити велики пакет патената, а да се пре тога не анализирају технологије и простор патената у односу на конкуренцију и евентуално снага истих. Постављена је следећа питање на која је потребно наћи одговор: Да ли компанија „Google“ може да пружи одговарајућу заштиту својим партнерима који у својим телефонима (енгл *Smart-phones*) искористе Андроид оперативни систем уколико нека од

конкурентских компанија покрене судски спор? У циљу проналажења одговора на постављено питање први у низу неопходних корака је претрага свих судских спорова који се тренутно воде а односе се на производе који су засновани на Андроид оперативном систему (енгл “*Smart-phones*”). Примат на овом тржишту су последњих година почели да преузимају уређаји засновани на Андроид оперативном систему и самим тим да угрожавају већ добро етаблиране произвођаче као што су “*Apple*”, “*Microsoft*” и “*RIM*” који продају уређаје засноване на властитим оперативним системима: „*iOS*“, „*Windows Mobile*“ итд.

6.4.2.1 Патенти који су укључени у спорове против Андроид-а

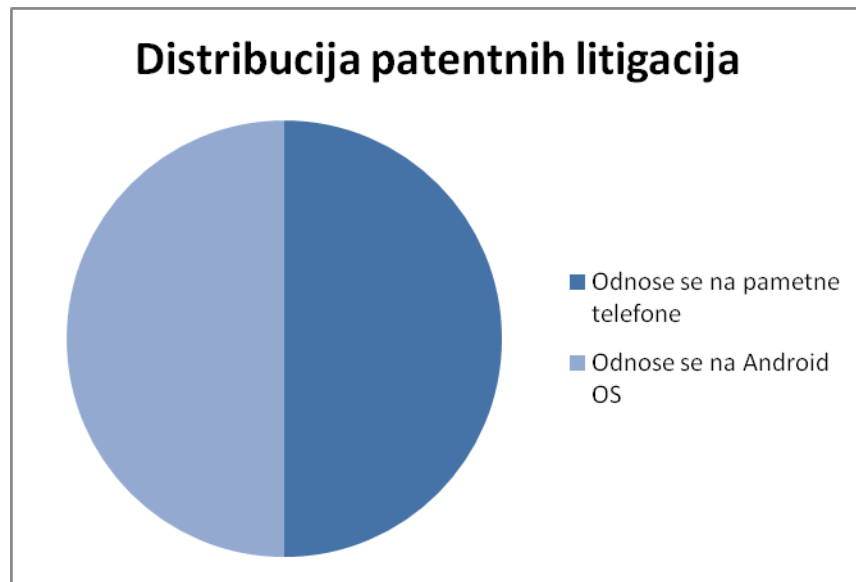
Комплетан преглед спорова који се тренутно воде, а усмерени су на Андроид ОС дат је у табели

Редни број	Тужилац	Број патента	Применљив на Андроид ОС
1.	Microsoft	US5,778,372	Да
2.	Microsoft	US6,339,780	Не
3.	Microsoft	US5,889,522	Да
4.	Microsoft	US6,891,551	Не
5.	Microsoft	US6,957,233	Да
6.	Microsoft	US5,579,517	Да
7.	Microsoft	US5,758,352	Да
8.	Microsoft	US6,621,746	Да
9.	Microsoft	US6,826,762	Не
10.	Microsoft	US6,909,910	Не
11.	Microsoft	US7,644,376	Не
12.	Microsoft	US5,664,133	Да
13.	Microsoft	US6,578,054	Да
14.	Microsoft	US6,370,566	Не
15.	Apple	US5,946,647	Да
16.	Apple	US6,343,263	Да
17.	Apple	US7,844,915	Не
18.	Apple	US7,469,381	Не
19.	Apple	US7,084,859	Не
20.	Apple	US7,920,129	Не
21.	Apple	US6,956,564	Не
22.	Apple	US7,362,331	Да
23.	Apple	US7,479,949	Не
24.	Apple	US7,657,849	Не
25.	Apple	US5,920,726	Не
26.	Apple	US7,633,076	Не
27.	Apple	US5,848,105	Не

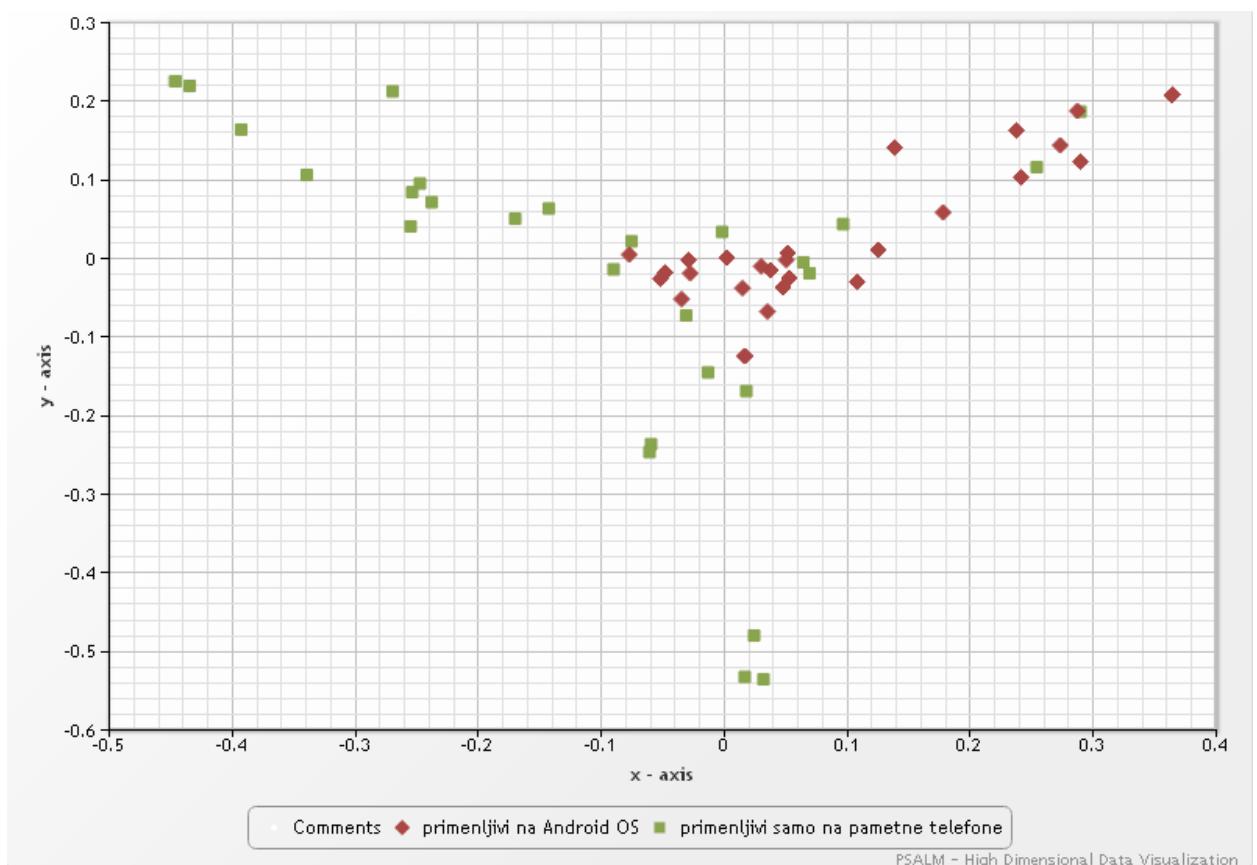
28.	Apple	US7,383,453	Да
29.	Apple	US5,455,599	Да
30.	Apple	US6,424,354	Да
31.	Apple	US5,481,721	Да
32.	Apple	US5,519,867	Да
33.	Apple	US6,275,983	Да
34.	Apple	US5,566,337	Да
35.	Apple	US5,929,852	Да
36.	Apple	US5,969,705	Да
37.	Apple	US5,915,131	Да
38.	Apple	RE39,486	Да
39.	Apple	US7,812,828	Не
40.	Apple	US7,669,134	Да
41.	Apple	US6,493,002	Да
42.	Apple	US7,853,891	Да
43.	Apple	US7,863,533	Не
44.	Oracle America, Inc	US6,125,447	Да
45.	Oracle America, Inc	US6,192,476	Не
46.	Oracle America, Inc	US5,966,702	Да
47.	Oracle America, Inc	US7,426,720	Да
48.	Oracle America, Inc	RE38,104	Да
49.	Oracle America, Inc	US6,910,205	Да
50.	British	US6,151,309	Не
51.	British	US6,169,515	Не
52.	British	US6,397,040	Не
53.	British	US6,578,079	Не
54.	British	US6,650,284	Не
55.	British	US6,826,598	Не

Табела 15 Приказ парница усмерених против Андроид ОС-а

Од укупно 55 патената приказаних у Табела 15 половина је применљива на сам Андроид ОС. То чини 50% од укупних парница које се тренутно воде (Слика 37). Ако ово прикажемо алатом ПСАЛМ, поред чињенице да су патенти подједнако груписани у две групе, видеће се и њихов распоред у 2Д простору технологија, што даје бољи увид и у њихов међусобни однос, тј. технолошку сличност (Слика 38).



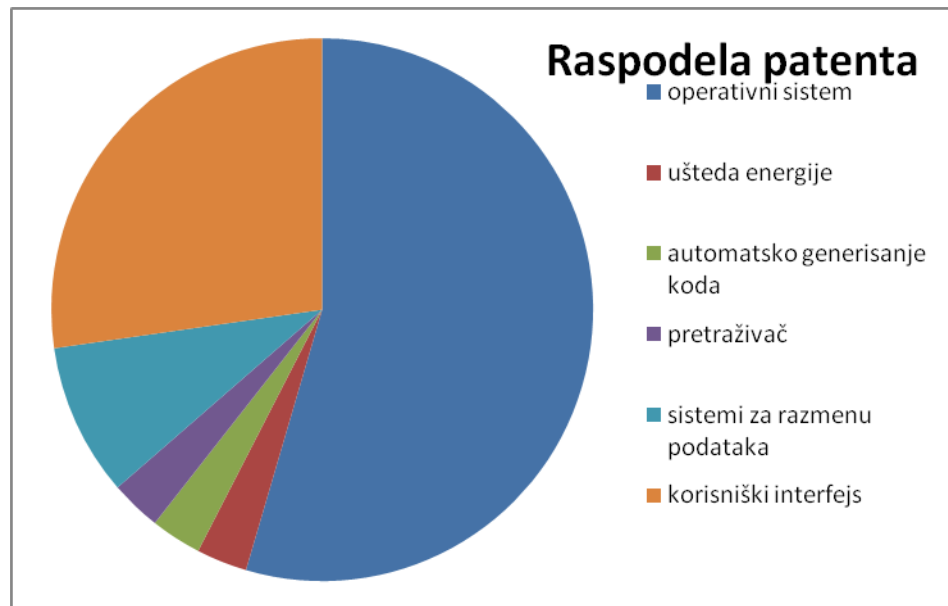
Слика 37 Дистрибуција патентних спорова



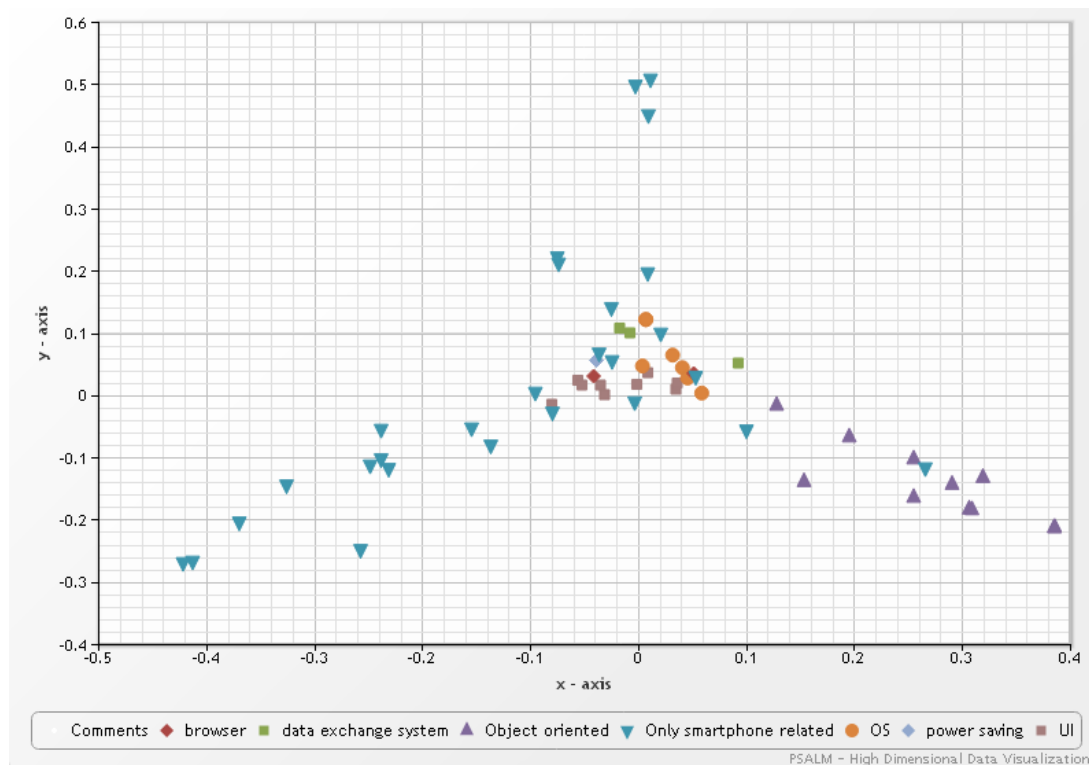
Слика 38 Дистрибуција патентних спорова приказана ПСАЈМ алатом

6.4.2.2 Класификација патената по технологији на коју се односе

Гледано по технологијама које су заступљене у патентима који се односе на Андроид ОС, највише је патената који се односе на сам оперативни систем, затим следе патенти који се односе на спрегу са корисником итд (Слика 39). Слика 40 приказује припадност патената одређеној технологији помоћу алата ПСАЈМ.



Слика 39 Расподела патената који се односе на Андроид ОС

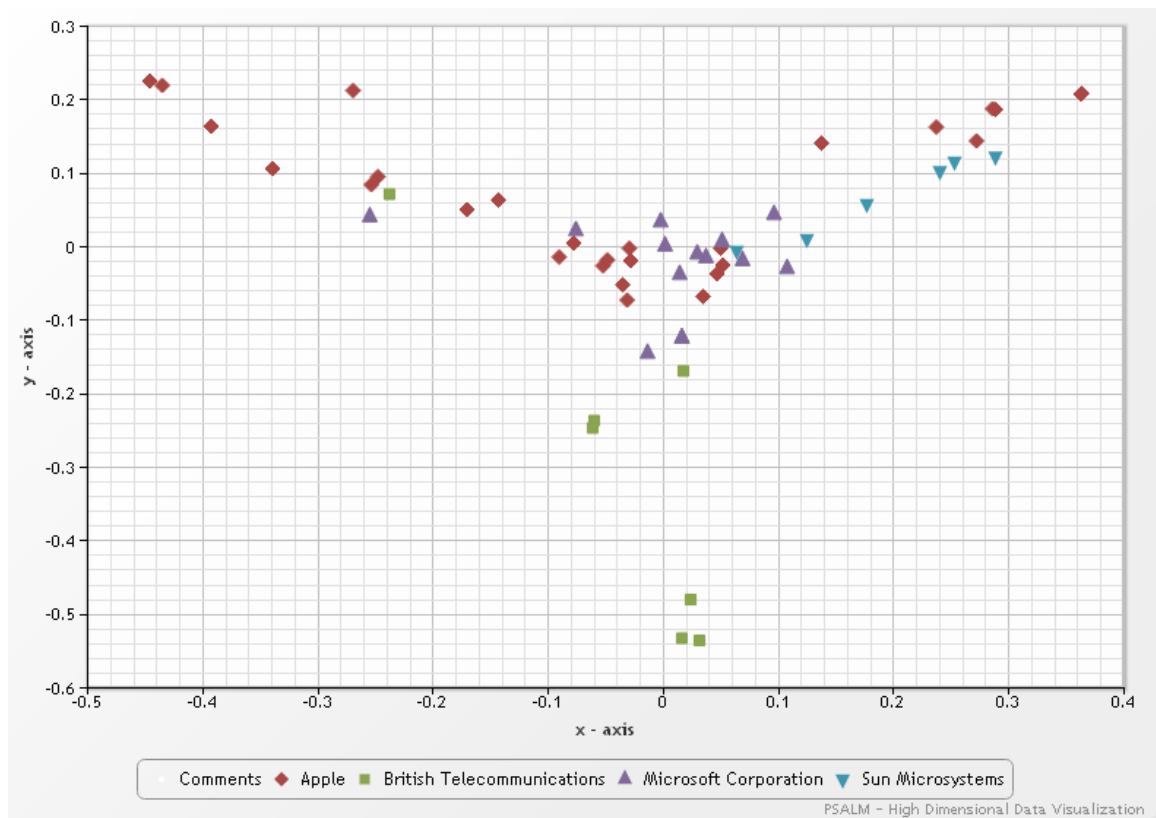


Слика 40 Расподела патената који се односе на Андроид ОС по технологији

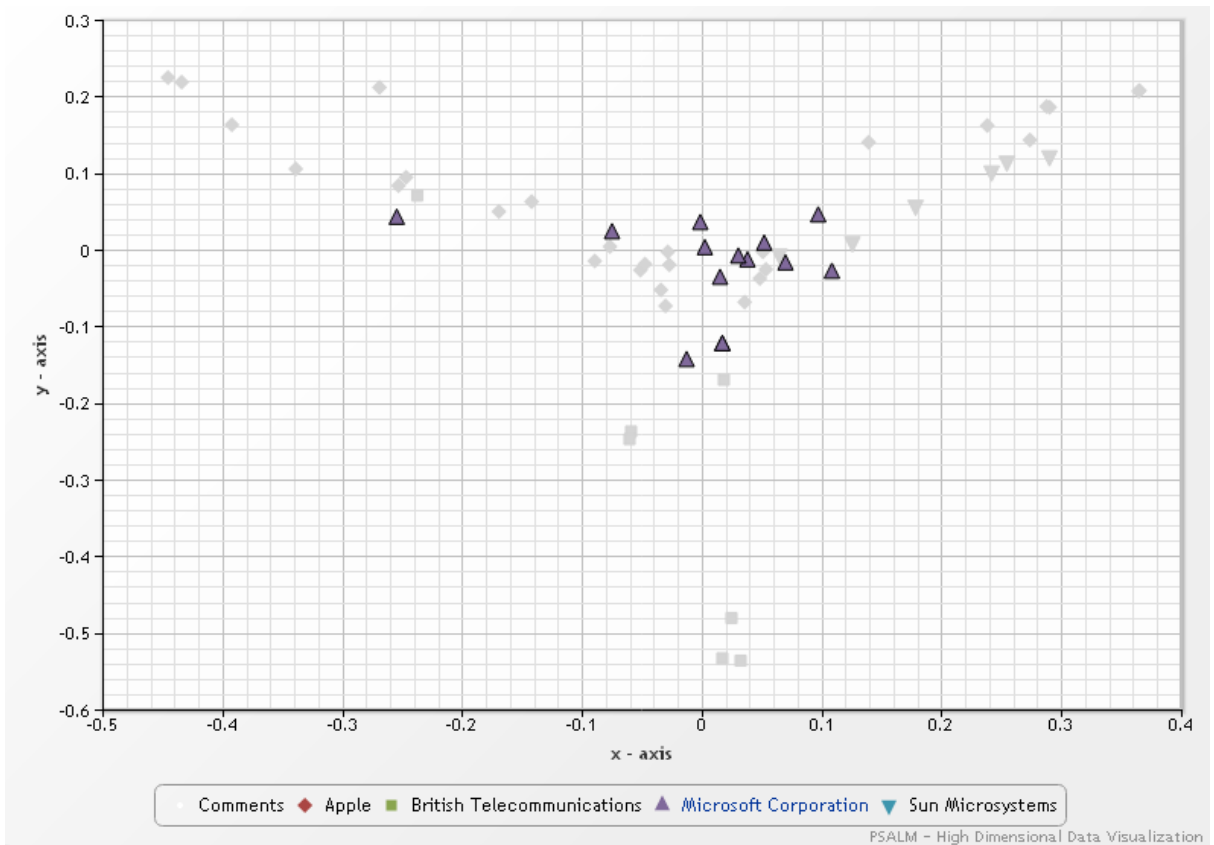
6.4.2.3 Класификација патената по власнику патентног права

Још један од могућих начина анализе истог скупа патената је и њихов приказ у простору технологија али груписаних по власништву патентног права (*Assignee*). Слика 41 представља визуализацију таквог груписања. Са дате слике може се видети да су патенти прилично добро груписани у овако формиране групе, из чега се може извући закључак да свака компанија напада компаније у спору из другог угла посматрано са становишта

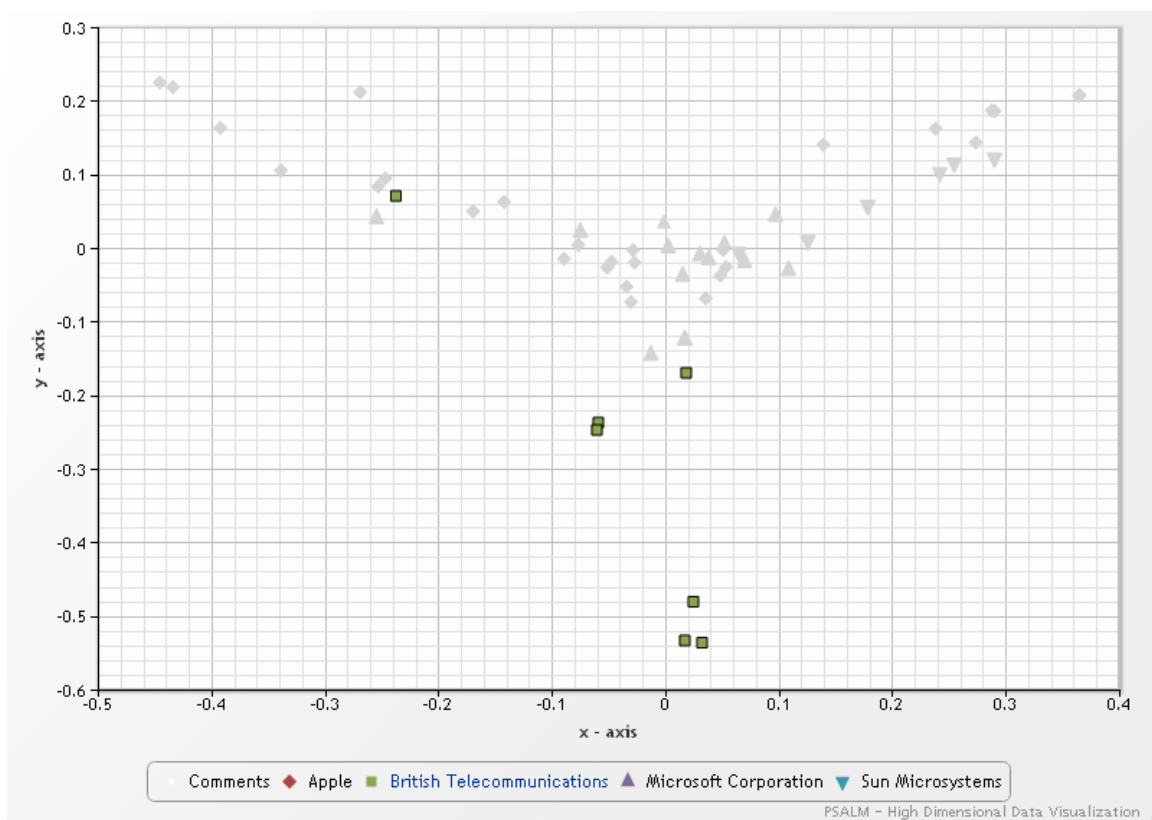
технологија. Слика 42 и Слика 43 приказују издвојене групе компанија „*Microsoft*“ и „*British Telecommunications*“ у односу на остатак, респективно.



Слика 41 Расподела патената у спору приказана по власницима патентних права

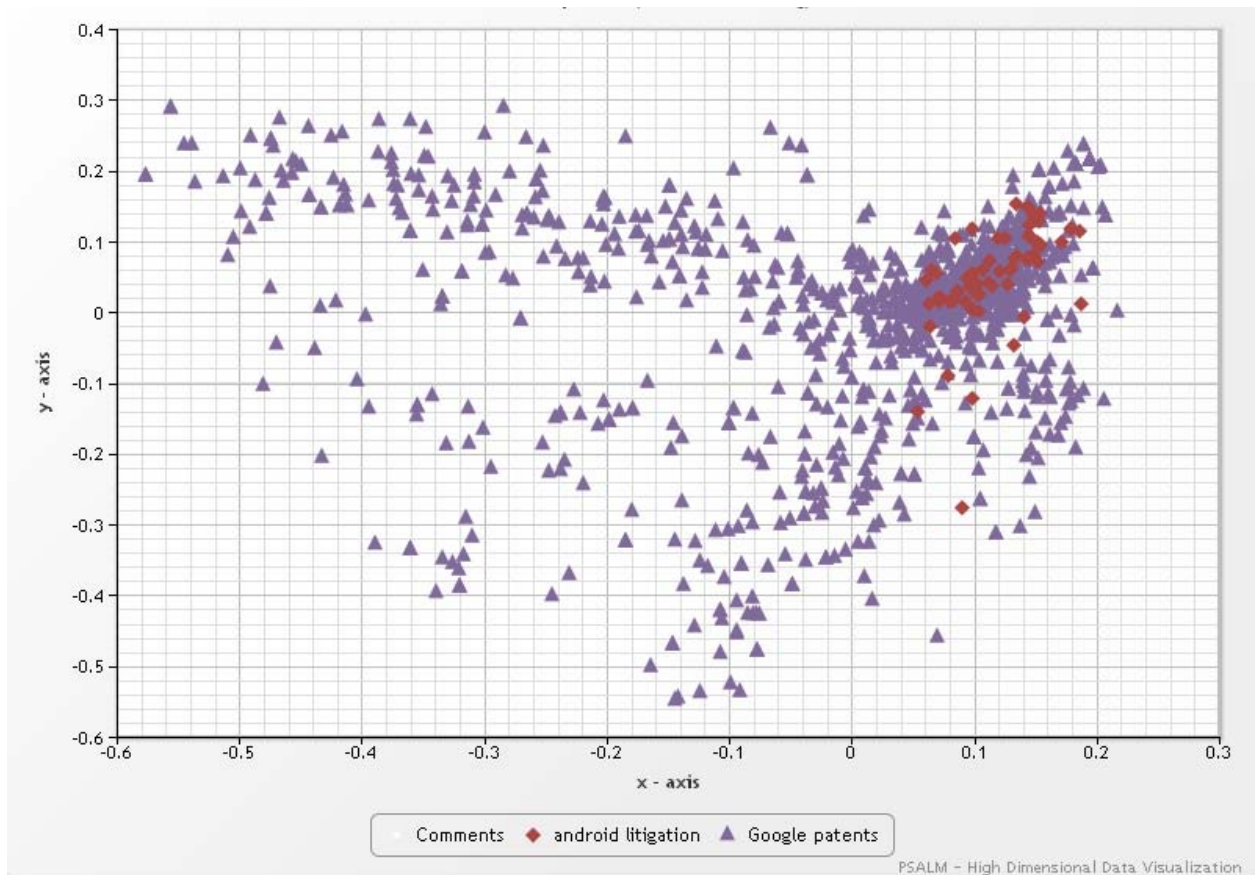


Слика 42 Расподела патената по власнику патентног права - *Microsoft*



Слика 43 Расподела патената по власнику патентног права – *BT*

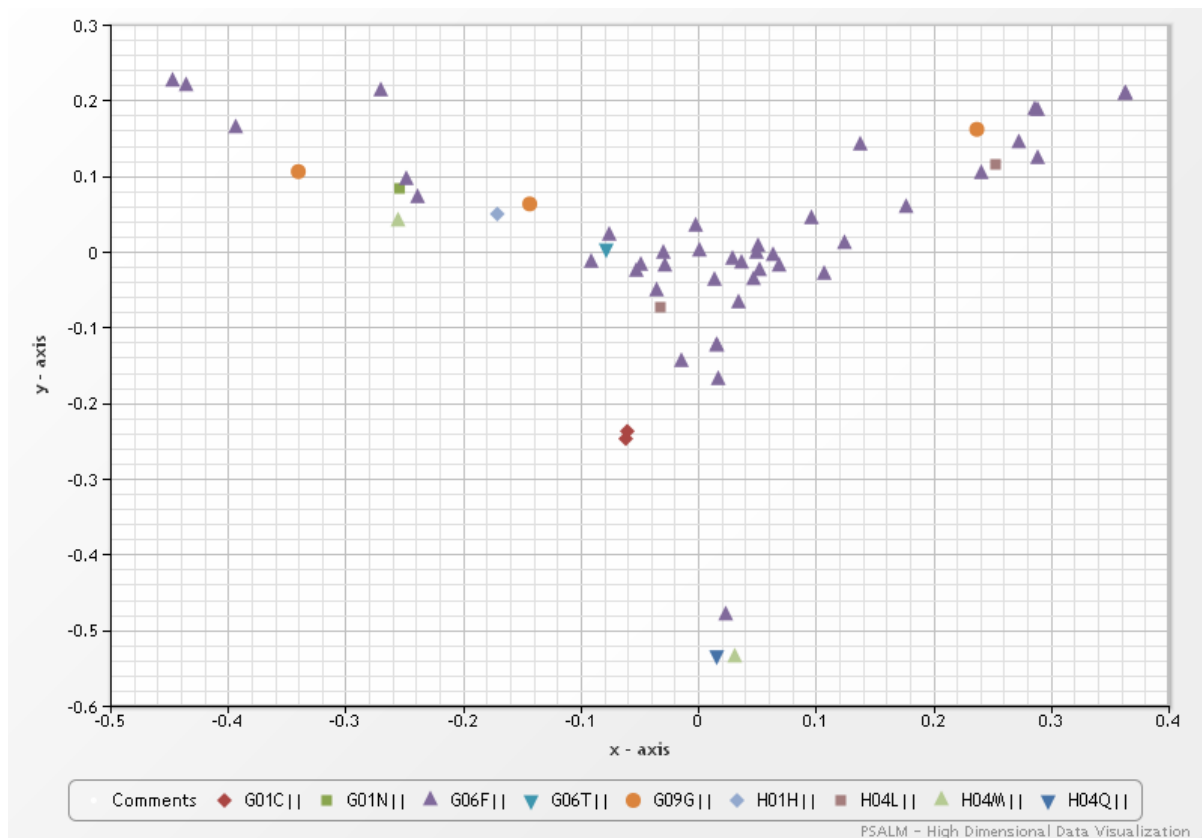
Претрагом патената који су у власништву компаније „*Google*” утврђено је да она у свом власништву има свега око 1000 одобрених патената. Слика 44 приказује дате патенте у 2Д простору у односу на патенте у спору. Са слике би се могло закључити да компанија „*Google*” има доста добру расподелу патената у односу на патенте у спору.



Слика 44 Литигациони патенти и патенти компаније „*Google*“

6.4.2.4 Класификација патената методом „груписање по кодовима“

Ако се дати скуп патената погледа из још једног угла – најзаступљенијим кодовима, исти се могу груписати методом „груписање по кодовима“. Слика 45 приказује резултат примене методе груписања. Може се уочити да је убедљиво најзаступљенија група „*G06F*“ (која представља дигиталне процесоре за обраду података) чак 42 патента од укупно 55 је сврстано у ову групу.



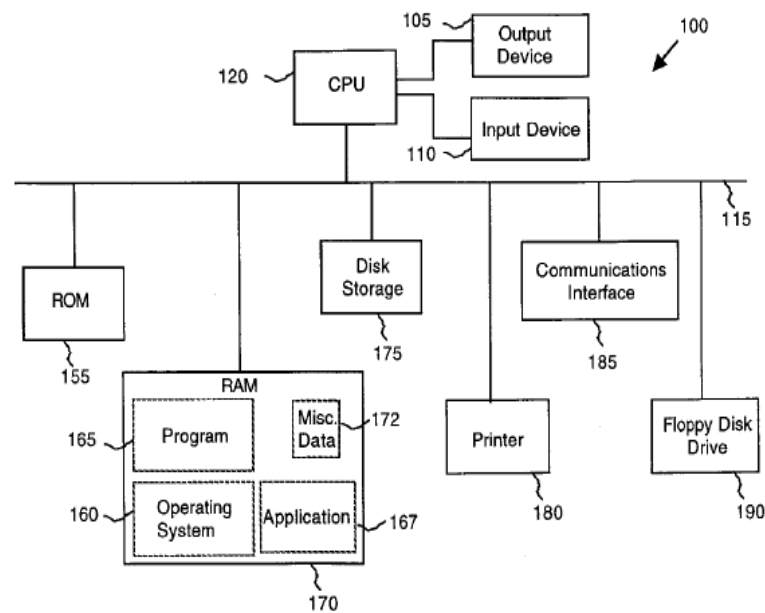
Слика 45 Груписање по кодovima

6.4.2.5 Анализа кључних патената

Постоје патенти у групи патената у спору на основу којих су компаније које туже већ добиле неке прелиминарне или коначне судске пресуде у своју корист. У овом поглављу ће бити анализирана три кључна патента у групи од 55.

Врло је значајан спор између компанија “Apple” и “HTC”. У првој пресуди која је изречена тражи се од компаније “HTC” да почевши од 12. Марта 2012. године заустави увоз у Сједињене Америчке Државе свих уређаја који нарушавају патент US5946647.

6.4.2.5.1 Патент [US5,946,647](#)



Слика 46 Патент US5,946,647

Патент 5,946,647 је амерички патент у власништву компаније “Apple” због којег је компанији “HTC” забрањен увоз производа на тржиште Сједињених Америчких Држава. Наиме, ИТЦ комисија је пресудила да компанија “HTC” (тј. неки њени производи који се продају на тржишту Сједињених Америчких Држава) нарушавају патентне захтеве број 1 и 8. Патентни захтеви 1 и 8 се односе на Андроид оперативни систем и то конкретно на проналажење одређених структура у подацима и затим спровођење одређених акција над тим структурама.

Захтев број 1 говори о систему заснованом на рачунару који је намењен откривању структура у подацима и спровођењу одређених акција над детектованим структурама:

„1. A computer-based system for detecting structures in data and performing actions on detected structures, comprising:

an input device for receiving data;

an output device for presenting the data;

a memory storing information including program routines including

an analyzer server for detecting structures in the data, and for linking actions to the detected structures;

a user interface enabling the selection of a detected structure and a linked action; and

an action processor for performing the selected action linked to the selected structure; and

a processing unit coupled to the input device, the output device, and the memory for controlling the execution of the program routines. “

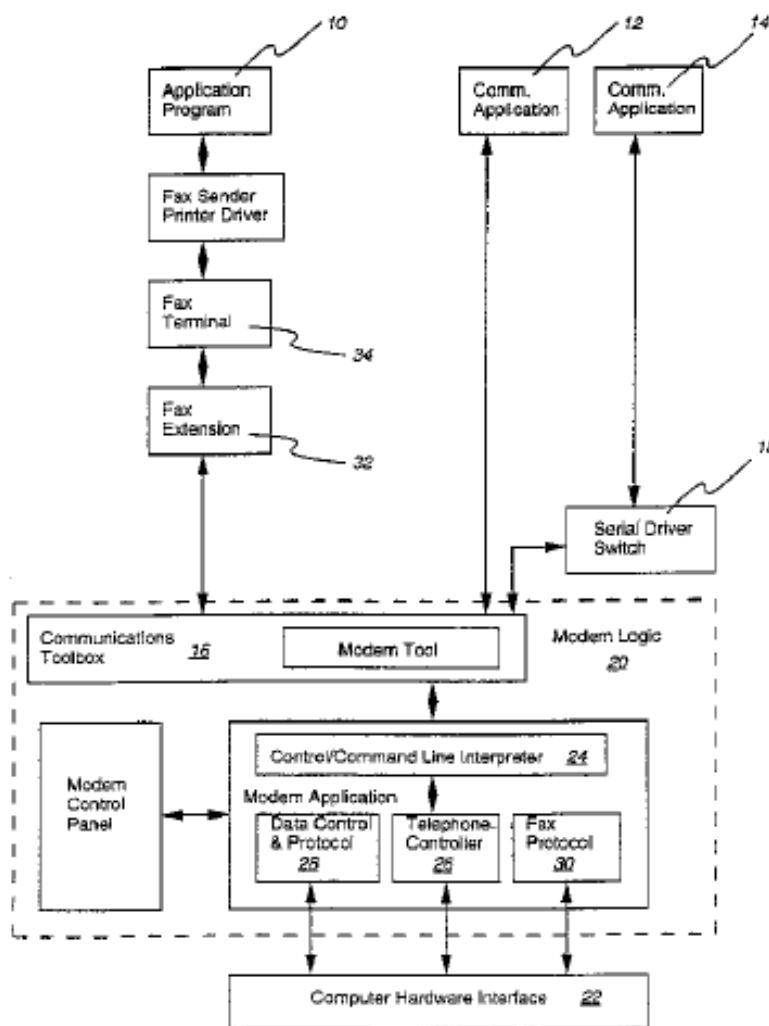
Захтев 8 говори о систему који је описан захтевом 1 и који је додатно одређен тиме што корисничка спрега може да нагласи откривене структуре.

„8. The system recited in claim 1, wherein the user interface highlights detected structures. “

Из описа ових захтева јасно је да се они не односе само на мобилне телефоне, већ се могу односити на било који уређај који има сличну функционалност, а заснован је на рачунару.

Други једнако битан патент којим компанија “Apple” прети да забрани дистрибуцију производа базираних на “Google” оперативном систему је патент „US6343263“.

6.4.2.5.2 Патент [US6,343,263](#)

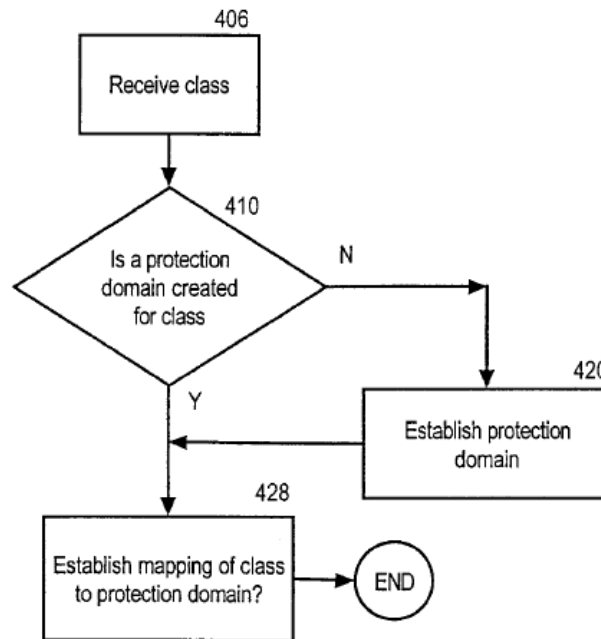


Слика 47 Патент US6,343,263

Патент се односи на обраду сигнала у реалном времену у случају серијског преноса података. Ово је чисто хардверски патент који има и различите референце на логичке компоненте у систему као што су програми, драјвери итд. Генерално, патент се односи на обезбеђивање апстрактних нивоа програмирања за апликације које се односе на обраду података у реалном времену. Ово је кључна компонента Андроид-а и може да има утицаја на цео Андроид-ов систем.

Суочена са тужбом која је против ње покренута и оваквим развојем догађаја компанија “HTC” је купила другу компанију под називом „S3 Graphics“ највероватније из једног јединог разлога: што је поменута компанија „S3 Graphics“ добила у своју корист прелиминарну пресуду против компаније “Apple” у патентном спору. Компанија “HTC” је донела овакву одлуку вероватно да би била у могућности да се судски поравна са компанијом “Apple”.

6.4.2.5.3 Патент [US6,125,447](#)



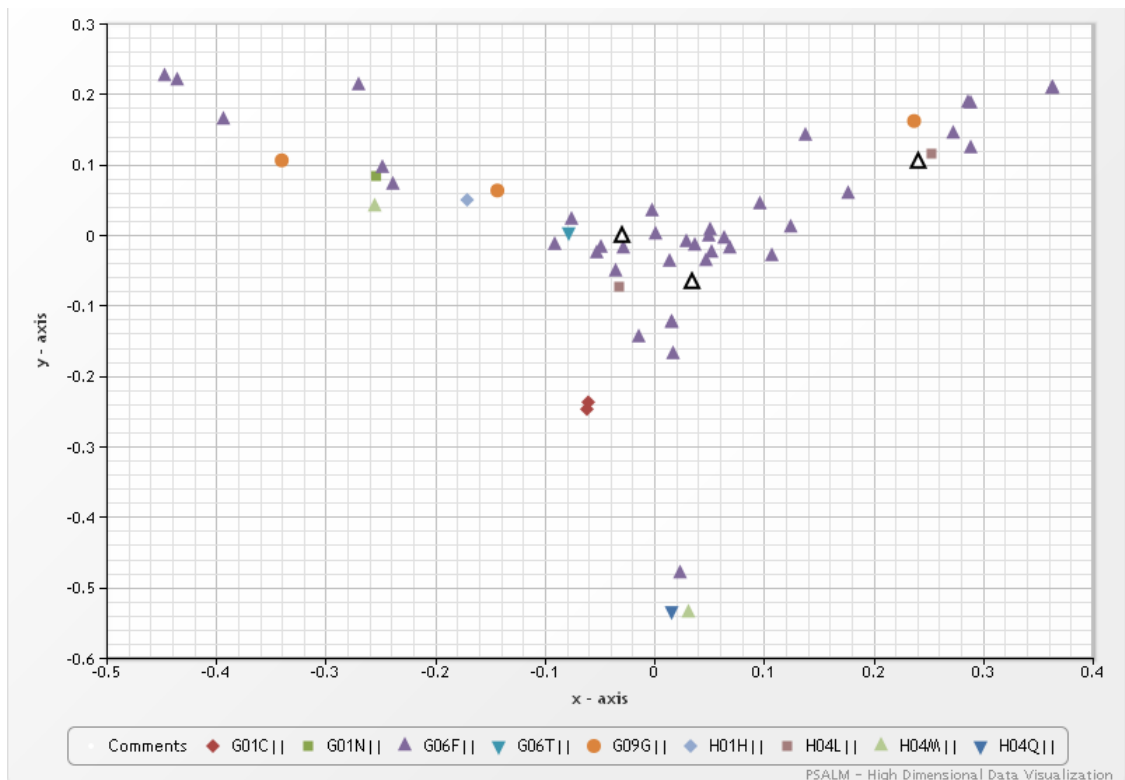
Слика 48 Патент US6,125,447

Патент US6,125,447 се односи на метод и уређај за примену и одржавање правила безбедности помоћу заштитних домена.

Захтев број један гласи:

*“A method for providing security, the method comprising the steps of:
 establishing one or more protection domains, wherein a protection domain is associated with zero or more permissions;
 establishing an association between said one or more protection domains and one or more classes of one or more objects; and
 determining whether an action requested by a particular object is permitted based on said association between said one or more protection domains and said one or more classes.”*

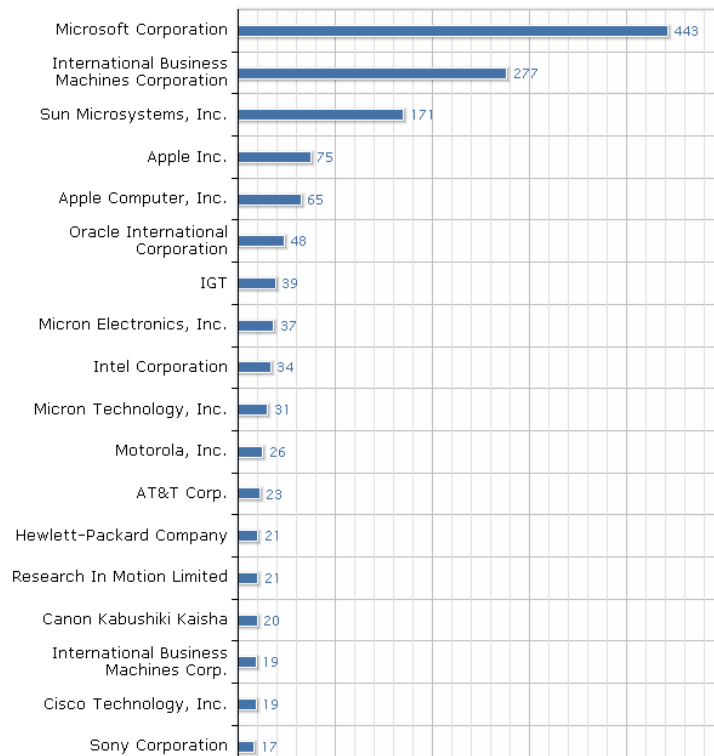
Слика 49 показује да и ова три, можда и најважнија патента у тренутним споровима, припадају класи „G06F“ као и већина других. Из приложеног би се могао извући закључак да је ова класа заједнички именилац за скоро све патенте у спору.



Слика 49 Три анализирана патента у 2Д простору

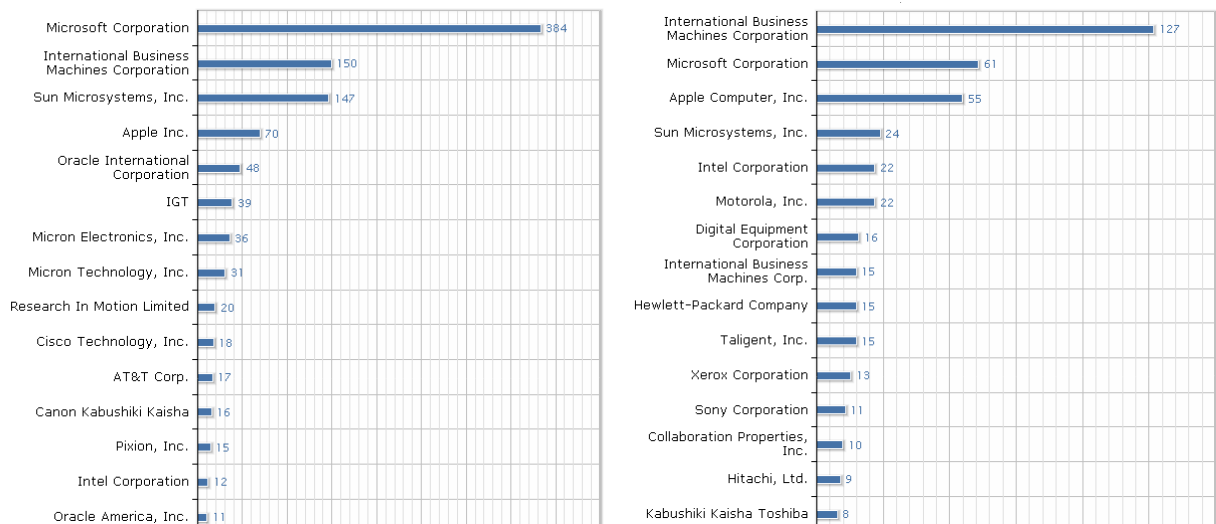
6.4.2.6 Патенти у власништву компаније „Motorola“

Да би се испитало да ли нека компанија искористила своју техничку спрему за грађење кључне технологије око неког производа, могу се испитати цитирани патенти од патената које је та компанија поднела током периода развоја датог производа. Разлог због кога има смисла посматрати цитиране патенте је тај што се очекује да што више компанија цитира своје претходне патенте, то више она гради своју базу знања, посебно ако је реч о новом технолошком пољу. Поред тога анализом цитираности се може доћи и до компанија које се баве сличним областима технологије. Предности ове анализе су приказане у овом поглављу како би се утврдило да ли је одлука компаније „Google“ да купи патенте компаније „Motorola“ била исправна. Користиле су се предности програмске подршке да се анализирају цитирани патенти и патенти који цитирају 55 патената у спору. Слика 50 приказује све цитате патената који су у споровима. Могло би се рећи да је компанија „Motorola“ заиста јака у области у којој се воде спорови јер има укупно 26 патената који цитирају неке од патената у спору или ови патенти цитирају њих, што је сврстава у првих 10 компанија које имају патенте из врло блиске области технике којом се баве патенти у спору.



Слика 50 Сви цитати патената у спору

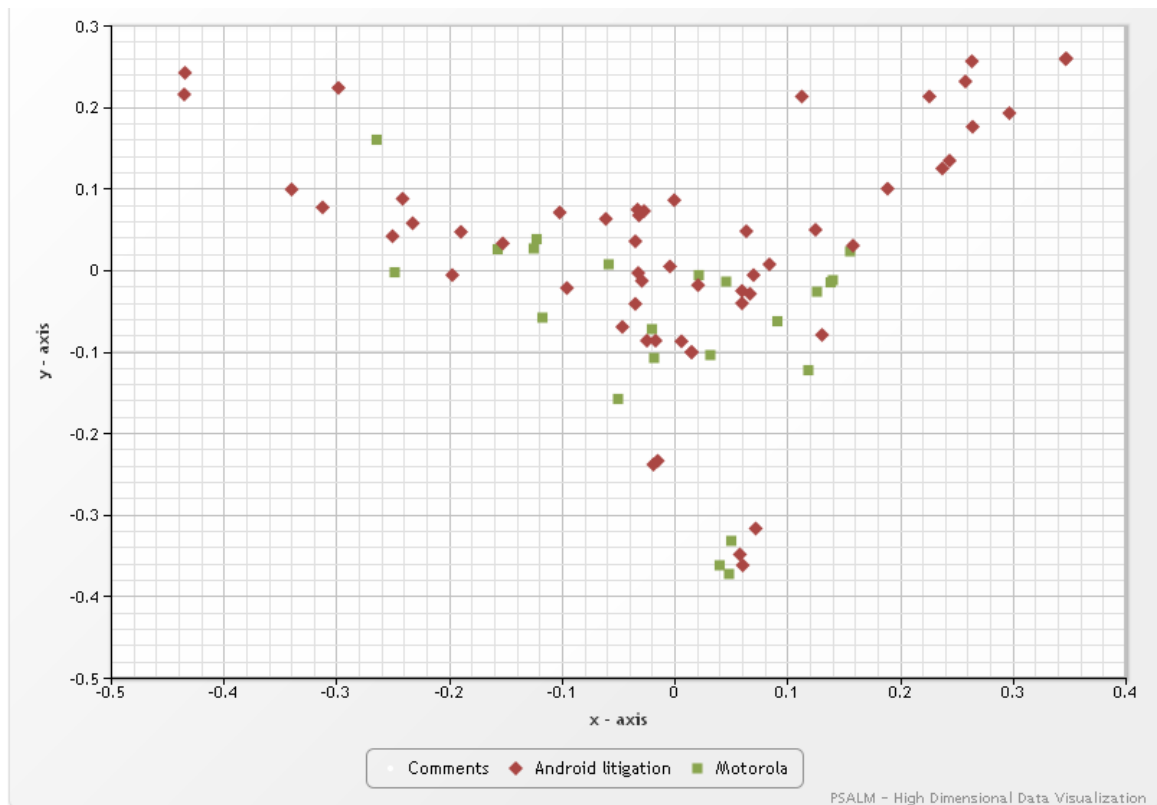
Међутим ако се цитати погледају мало детаљније и раздвоје се патенти који цитирају патенте од оних који су цитирани добијају се следећи дијаграми.



Слика 51 Патенти који цитирају (лево) и патенти који су цитирани (десно)

Са дијаграма се види да компанија „Motorola“ има 22 патента у листи цитираних патената (*backward citations*), што значи да преостају само четири патента који цитирају патенте у спору (*forward citations*). Тих 22 патента су старији од патената у спору и могу представљати основ за евентуалну против-тужбу. 22 патента гледано квантитативно не уливају велику наду да би могли представљати основ за одбрану у односу на 55

литигационих патената. Слика 52 показује однос ова 22 патента у 2Д простору заједно са патентима у спору.



Слика 52 Патенти компаније „*Motorola*“ у односу на литигационе

Са последњег дијаграма види се да патенти компаније „*Motorola*“ јесу испреплетени са патентима у судским споровима, али да постоје делови простора у којима их уопште нема.

6.4.2.7 Закључак

Гледано чисто квантитативно, исувише мали портфолио да би компанија могла да одговори на све претње својих конкурената. „*Google*“ тренутно нема јак портфолио да може да брани Андроид ОС нити се преузимањем оперативног система стичу права над интелектуалном својином. Тако су компаније препуштене саме себи да се бране пред евентуалним парницама које би против њих биле покренуте.

Анализом цитираности литигационих патената утврђено је да су патенти компаније „*Motorola*“ релативно добро распоређени у 2Д простору у односу на патенте у споровима, те се са те стране не може рећи да је одлука о куповини њеног патентног портфолија била погрешна. Са друге стране компанија „*Motorola*“ нема патенте довољно блиске патентима у споровима, тако да ће компанија „*Google*“ по свему судећи и поред ове куповине морати да обави још неколико значајних куповина на тржишту патената како би могла да се супротстави компанијама које је нападају.

7. Закључак

У овом раду представљено је једно решење аутоматске обраде патентних докумената које је развијено са циљем унапређивања руковања патентним портфолијима компанија. Употребна вредност је показана на реалним примерима, а практични резултати су верификовани и субјективним и објективним методама. Почевши од перформанси интернет робота, преко илустрације расподела патената у 2Д простору које се је показало задовољавајуће резултате, затим властитих метода за класификацију које су показале добре резултате, па до визуализације која је омогућила да се све могућности решења приближе кориснику.

Са становишта квантитативног груписања патената у групе које немају пуно издвојених тачака у односу на број тачака у групи, обе властите методе су показале добре резултате јер се тачност креће у интервалу од 90 до 100 процената. Визуализацијом су такође потврђени резултати добијени математичким путем.

Овај рад дао је доприносе тренутном стању технике и техничких решења у области аутоматске обраде патентних докумената. Даља могућност развоја програмске подршке огледа се, између осталог, на проширивање претраге патентних докумената на следеће две најзначајније базе патената у свету: Европског патентног завода (енгл. *EPO*) и Светског завода за интелектуалну својину (енгл. *WIPO*). Једно од неопходних проширења интернет робота је претрага база података како би се поред оригиналног власника патентног права аутоматски проналазио и тренутни власник патентног права. Такође, постоји могућност проширења функционалности и повећања тачности програмске подршке додавањем, поред лексичке, семантичке обраде текста патентног документа која би могла водити чак до идентификовања евентуалног нарушавања патента. Оно што би сигурно у значајној мери допринело побољшању тренутне функционалности решења је и проширење функције вредновања патената и проширити истраживање и на случај како утврдити

вредност патента у односу на чињеницу да ли је патентно решење нарушено? У том случају би базу података требало проширити и са другом доступном литературом као што су научни часописи и други извори информација на основу којих би се могли проценити и пратити правци кретања развоја у компанијама.

8. Литература

- [1] Hyunseok Park, Janghyeok Yoon, Kwangsoo Kim *Identifying patent infringement using SAO based semantic technological similarities*, Scientometrics DOI 10.1007/s11192-011-0522-7, May 2011.
- [2] Dragan Kukolj, *SISTEMI ZASNOVANI NA RAČUNARSKOJ INTELIGENCIJI*, monografija, FTN izdavaštvo, 2007, ISBN 978-86-7892-068-4
- [3] Flora S. Tsai, *A visualization metric for dimensionality reduction*, Expert Systems with Applications 39 (2012) 1747–1752, Elsevier 2011
- [4] S. Haykin, *Neural Networks*, Macmillan Publ. Co, New York, 1994
- [5] R.P. Lippmann, *An Introduction to Computing with Neural Nets*, IEEE ASSP Magazine, April pp. 4-22, 1987
- [6] Jinlong Wang, Shunyao Wu, Gang Li, Zhe Wei, *Integrating Instance-level and Attribute-level Knowledge into Document Clustering*, DOI: 10.2298/CSIS100906003W, ComSIS Vol. 8, No. 3, June 2011
- [7] Chulhyun Kim, Hakyeon Lee, Hyeonju Seol, Changyong Lee, *Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach*, Expert Systems with Applications 38 (2011) 12559–12564, Elsevier 2011
- [8] Tad H. Hogg, Rajan M. Lukose „*Robust clustering of web documents*“, US patent 6654743, Nov 13, 2000
- [9] Yu Zhao, Jianqiang Li „*METHOD AND SYSTEM FOR WEB DOCUMENT CLUSTERING*“ US patent application 2009/0070366 A1, Sep 11, 2008
- [10] <http://www.matheo-software.com/en/products/matheo-patent.html>
- [11] <http://www.patentinsightpro.com/product.html>

-
- [12] https://thomsonipmanagement.com/docs/default-document-library/thomson_ip_manager_fs_0611.pdf
- [13] https://thomsonipmanagement.com/docs/default-document-library/process_architect_fs_0611D70693DA98AA.pdf
- [14] https://thomsonipmanagement.com/docs/default-document-library/data_validation_fs_0611E4348C830574.pdf
- [15] <https://www.cobaltip.com/>
- [16] PatVal-EU, *The value of the European patents: evidence from a survey of European inventors*. Final Report of the PatVal-EU Project, DG Science & Technology, European Commission, Contract N. HPV2-CT-2001-00013, 2005. Brussels
- [17] Jean O. Lanjouw, Ariel Pakes, Jonathan Putnam *How to count patents and value intellectual property: uses of patent renewal and application data*, National Bureau of Economic Research, Cambridge, September 1996.
- [18] Hughes A, Mina A. *The impact of the patent system on SMEs*, Centre for Business Research, University of Cambridge, 2010. Working Paper No.411, www.cbr.cam.ac.uk/pdf/WP411.pdf
- [19] Ruotsalainen, L. (2008). *Data mining tools for technology and competitive intelligence*. Espoo 2008. VTT Tiedotteita – Research Notes 2451
- [20] Feldman R. Sanger J. *The text mining handbook – advanced approaches in analyzing unstructured data*, Cambridge University Press: New York. 2007
- [21] H.C. Wu, R.W.P. Luk, K.F. Wong, K.L. Kwok (2008). "Interpreting tf-idf term weights as making relevance decisions". *ACM Transactions on Information Systems* 26 (3): 1–37.
- [22] T. Cox, M. Cox, (1994) *Multidimensional Scaling*, London: Chapman & Hall, 1994.
- [23] Nock, R. and Nielsen, F. (2006) "On Weighting Clustering", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28 (8), 1–13
- [24] Keim, D. A. 2002. *Information Visualization and Visual Data Mining*. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 8, No. 1, January-March.
- [25] D. Kukolj, Z. Tekic, Lj. Nikolic, Z. Panjkov, M. Pokric, M. Drazic, M. Vitas, D. Nemet, *Comparison of Algorithms for Patent Documents Clusterization*, Mipro 2012

-
- [26] Ljubiša Nikolić, Dragan Kukolj, Maja Pokrić, Miroslava Dražić, Marina Vučković, Milana Vitas *INTERNET ROBOT – PREUZIMANJE PODATAKA O PATENTIMA SA INTERNET STRANICE POZNATE STRUKTURE*, Etran, 2012
- [27] Salton G, Buckley C. *Term-weighting approaches in automatic text retrieval*, Information Processing & Management 1998: 24 (5): 513–523
- [28] Zeljko Tekic, Dragan Kukolj, Ljubisa Nikolic, Maja Pokric, Miroslava Drazic, Milana Vitas: *SMEs, Patent Data and New Tool for Business Intelligence*, Proceedings of International Conference for Entrepreneurship, Innovation and Regional Development ICEIRD 2012
- [29] <http://www.ft.com/intl/cms/s/2/de24f970-f8d0-11e0-a5f7-00144feab49a.html#axzz1caYTb5ix>
- [30] <http://www.trefis.com/stock/aapl/articles/106425/apples-victory-over-motorola-in-essential-patent-case-is-huge/2012-03-01?from=email%3Anotd>