

Lappeenrannan teknillinen yliopisto
Tietotekniikan osasto
Kandidaatintyö

PATSTAT -tietokannan hyödyntäminen tutkimuskäytössä

Työn ohjaaja ja tarkastaja: DI Matti Karvonen

Lappeenranta, 8.12.2009

Juha Kortelainen
Korpraalinkuja 1 as 103
53810 Lappeenranta
Puh: 040-7596725

TIIVISTELMÄ

Lappeenrannan teknillinen yliopisto
Tietotekniikan osasto

Juha Kortelainen

Patstat-tietokannan hyödyntäminen tutkimuskäytössä

Kandidaatintyö

2009

34 sivua, 8 kuvaa, 2 taulukkoa

Tarkastaja: DI Matti Karvonen

Hakusanat: patenti, tietokanta, Patstat

Keywords: patent, database, Patstat

Patenttitiedot sisältävät monenlaista tietoa joka on julkisesti saatavilla. Patenttien tietoja tutkimalla voidaan selvittää monenlaisia asioita, kuten yritysten tekemiä patenttimääriä, patenttiluokitusten jakautumista ja patenttien maantieteellistä jakaumaa. Vielä 1980-luvulla patenttitietojen analysoiminen suuremmissa määrin oli vaikeaa, sillä tietokoneiden laskentateho oli riittämätön. Nyt laskentatehon räjähdysmäisen kasvun ansiosta patenteja voidaan tutkia miljoonia kappaleita kerrallaan. Tässä työssä tutkitaan patenttitietokantoja, tietokantaserverin pystyttämistä ja tietojen hyödyntämistä jatkokäytössä.

ABSTRACT

Lappeenranta University of Technology
Department of Information Technology

Juha Kortelainen

Usage of Patstat database in research use

Bachelor's Thesis

2009

34 pages, 8 figures, 2 tables

Supervisor: M.Sc. Matti Karvonen

Keywords: patent, database, Patstat

Patent information contains many kind of information which is publicly available. By investigating patent information we can examine many kind of things, like amount of patents that a certain company has made, distribution of patent classifications and geographical distribution of patents. Even at 1980's the analysis of great numbers of patent information was difficult due the lack of computing power. Now due to explosive growth of computing power patents can be analyzed with millions of patents at one time. In this thesis we study patent databases, building a database server and using the patent information in research.

SISÄLLYSLUETTELO

LYHENTEET	2
1 JOHDANTO	3
1.1 Taustaa	3
1.2 Työn tavoitteet ja rajaukset	3
1.3 Työn rakenne.....	5
2 PATENTTITIEKANNAT TUTKIMUSKÄYTÖSSÄ.....	6
2.1 Sitaatioista.....	6
2.2 IPC-luokitus	9
2.3 Patentin käyttö.....	10
2.3.1 Patenttihakemuksen täyttäminen.....	10
2.4 Patentin hakuprosessi	10
2.5 Miksi patenti-informaatiota	12
2.5.1 Mitä informaatiota patenttidokumentti sisältää?.....	12
2.5.2 Mistä patenttitietoa voi löytää?	12
3 PATSTAT-tietokanta	14
3.1 Tekninen toteutus.....	14
3.1.1 phpMyAdmin	17
3.1.2 RAID	17
3.1.3 Tietokantamoottoreista.....	17
3.1.4 Indeksointi.....	18
3.1.5 Esille nousseet ongelmat Patstat-tietokannassa	18
4 TIETOKANNAN HYÖDYNTÄMINEN TUTKIMUSKÄYTÖSSÄ.....	21
4.1 Yrityksen nimen perusteella.....	21
4.1.1 Sitaatioista.....	24
4.2 Patenttiluokitusten avulla.....	25
5 JOHTOPÄÄTÖKSET.....	27

LYHENTEET

VPN	Virtual Private Networking
CSV	Comma Separated Values
WIPO	World Intellectual Property Organization
GHz	Gigahertsi
IPC	International Patent Classification
RAM	Random Access Memory
SCSI	Small Computer System Interface
RPM	Revolutions Per Minute
SAS	Serial Attached SCSI
SQL	Structured Query Language
OECD	Organisation of Economic Co-operation and Development
RFID	Radio Frequency Identification

1 JOHDANTO

1.1 Taustaa

Tämä kandidaatintyö liittyy patenttitietokannan käyttöönottoon Lappeenrannan teknillisen yliopiston (LUT) tiloissa. Yliopisto ei hyödynnä tällä hetkellä aktiivisesti tietokantaa tai muuta tietolähdettä josta saataisiin helposti ja kätevästi erilaista tietoa patenteista, joten tarve tällaiselle on olemassa. Tiedon saamisen lisäksi tietokantaa voitaisiin käyttää tilastolliseksi hyväksi suoritettaessa tutkimusta esimerkiksi vertaillen eri yritysten patenttiluokituksia.

Internetissä on olemassa myös patenttitietokantoja joita kuka tahansa voi hyödyntää ilmaiseksi. Näitä ovat mm. Espacenet, US-patent sekä freepatentsonline. Espacenet on maailmanlaajuinen patenttihakemisto, US-patent sisältää USA:n patenttien täydet tekstit ja freepatentsonline sisältää voimassaolevat US-patentit pdf-tiedostoina. [19, 20, 21]

Patenttitietoja tutkimalla ja analysoimalla saadaan kätevästi julkisesti saatavilla olevaa tietoa ja sitä voidaan linkittää myös muista tietokannoista saatuihin tietoihin. Patenttianalyysin avulla on mahdollista selvittää tietyn alan teknologiajohtajat ja kehityssuunnat sekä maantieteellinen jakauma.

1.2 Työn tavoitteet ja rajaukset

Työ liittyy LUT:n tuotantotalouden osaston jatko-opiskelijoiden sekä tutkijoiden tarpeisiin päästä hyödyntämään paremmin patenttitietokannasta saatavaa tietoa. Tällä hetkellä LUT:n tiloissa ei ole käytössä minkäänlaista patenttitietokantaa jonka tietoja voitaisiin hyödyntää tutkimuksissa.

Edellisessä kappaleessa mainituista tietokannoista yksittäisen tiedon hakeminen on helppoa ja tieto on myös tarkkaa. Tätä kautta ei kuitenkaan voida tehdä tutkimusta suuresta määrästä patenteja kerralla, joten ongelmana on sellaisen järjestelmän rakentaminen josta saisi suuria määriä tietoja kerralla analysoitavaksi.

Työn tavoitteena on LUT:n tutkijoille ja jatko-opiskelijoille suunnatun patenttitietokannan määrittäminen, pystytys sekä tietojen hyödyntämisen tutkiminen. Patenttiedot ovat yleensä hyvin tarkkaan määritettyjä, emmekä varsinaisesti käsittele kaikkea dataa, joka liittyy yksittäiseen patenttiin vaan keskitymme meneillään olevien tutkimuksien kannalta olennaisiin tietoihin.

Työssä selvitetään tarvittavat resurssit tietokannan teknistä toteutusta varten hyödyntämällä aiheeseen liittyvää kirjallisuutta sekä selvitetään kuinka patenttitietokannan tietoja voidaan hyödyntää tutkimuskäytössä. Itse tietokannan sisältämistä tiedoista ja tietokannan rakenteesta ei ole saatavilla paljoakaan tietoa valmiiksi vaan lähes pelkästään teknistä dokumentaatiota. Tästä johtuen tietokannan tekniset resurssit määritetään valmiiksi tarvittavalla tasolla varmistaaksemme riittävän suorituskyvyn monen yhtäaikaisen käyttäjän tietokantaympäristöön.

Painopisteenä työssä on myös patenttien ja niiden sitaatioiden tutkiminen keksintöjen synnyssä ja tiedonlevityksessä. Tutkimme mitä erilaiset sitaatiot tarkoittavat ja miten esimerkiksi sitaatioiden määriä voidaan tulkita.

Loppuosassa työtä perehdytään itse tietokannan tietojen käyttöönottoon ja tietojen hyödyntämiseen tutkimustyössä. Loppuosassa pyritään siis löytämään tarvittavat resurssit tietokannan tekniseen toteutukseen, tutustutaan tietokannan rakenteeseen ja mietitään kuinka tietokannan tietoja voidaan hyödyntää tutkimuksessa. Tietojen hyödyntämisen kohdalla ei käydä läpi SQL-kyselyitä, joilla tietoa haetaan tietokannasta.

Käymme myös läpi mahdollisesti esiin nousevat haasteet tai ongelmat. Ongelmiin voivat lukeutua mm. tietokannan monimutkainen rakenne, suuri koko tai liian suuret taulut, mikäli kyseessä on relaatiotietokanta. Haasteisiin voivat kuulua tietojen hakeminen tietokannasta, sen jalostamisen eri vaiheet siten että sitä voidaan hyödyntää tutkimuskäytössä ja epärelevantit tulokset tutkimuskohteen kannalta.

1.3 Työn rakenne

Työn alussa (luku 2) keskitytään kertomaan patenteista sekä patenttietokannoista yleisesti. Luvussa kerrotaan mitä patentit ovat, millaista tietoa niihin liittyy ja miten niitä voidaan hyödyntää tutkimuskäytössä.

Kolmannessa luvussa käymme läpi PATSTAT-tietokantaa. Luvussa pohdimme tietokannan rakennetta, sen asettamia vaatimuksia, tietokannan toteutusta kuten myös tietokantapalvelimen pystyttämistä. Perehdymme asioihin jotka tulee huomioida tarvittavan suorituskyvyn ja palveluiden turvaamiseksi ja millaisia vaihtoehtoja niille on mahdollisesti olemassa. Määritämme myös kuinka ja mitä apuohjelmia käyttäen tietoa saadaan haettua tietokannasta jatkokäsittelyä varten. Neljännessä luvussa käymme läpi konkreettisten esimerkkien avulla kaksi erilaista tietokantahakua, joiden tuloksia havainnollistamme kuvaajien avulla. Viides ja viimeinen luku käsittää katsauksen koko tehtyyn työhön ja toimii yhteenvetona ja johtopäätöksinä.

2 PATENTTITIETOKANNAT TUTKIMUSKÄYTÖSSÄ

Patenttitietokannat sisältävät valtavia määriä tietoa ja jokaiseen yksittäiseen patenttiin liittyy montaa eri tietoa. Kappaleessa käymme läpi patenttien tietokenttiä, patenttien hakuprosessin ja mihin patentti-informaatiota voidaan käyttää.

2.1 Sitaatioista

Patenttien "sitatiot" ovat viittauksia edellisiin töihin, jotka ovat oleellisia viittaavan, eli haettavan patentin kannalta. Sitaatioita voisi verrata tutkimuspapereistakin löytyviin viittauksiin muiden tekemiin tutkimuksiin [15]. Patenttitoimistojen tutkijat käyttävät näitä viittauksia hyväkseen tutkiessaan onko jokin tietty keksintö todellakin uusi. Viittaukset voidaan tehdä muihinkin asioihin kuin patenteihin, kuten tieteellisiin julkaisuihin, mutta tässä työssä viittauksilla tarkoitetaan vain toisia patenteja. [3]. Patenttien sitaatioiden sanotaan heijastavan patenttien laatua ja toimivan myös teknologisen tiedon indikaattoreina [14].

Patentteja on pitkään pidetty runsaana tiedonlähteenä jolla voidaan tutkia mm. innovatiivisuutta ja teknologian muuttumista. Patenttiedon käyttämisestä on yleisesti koettu olevan seuraavia hyötyjä:

- Jokainen patentti sisältää erittäin tarkkaa tietoa itse innovaatiosta, sen teknologisesta alueesta, keksijästä ja maantieteellisestä sijainnista.
- Patentit antavat väliaikaiset, monopoliaseman tapaiset oikeudet sillä ehdolla että patentin sisältämä tieto julkaistaan.
- Patenttidata sisältää sitaatioita edellisiin patenteihin ja kirjallisuuteen. Näiden sitaatioiden avulla voidaan selvittää keksintöjen, keksijöiden, yritysten yms. suhteita toisiinsa. Yksi tärkeimmistä sitaatiokäytöistä on myös tärkeiden patenttien selvittäminen, joita ovat siis sellaiset joihin on viitattu usein. [9].

Kaikkia keksintöjä ei kumminkaan patentoida, koska ne eivät esim. täytä tiettyjä vaatimuksia olemalla tarpeeksi uudenlaisia tai kaupallisesti hyödynnettävissä. [9]. Keksinnöiksi ei katsota eikä

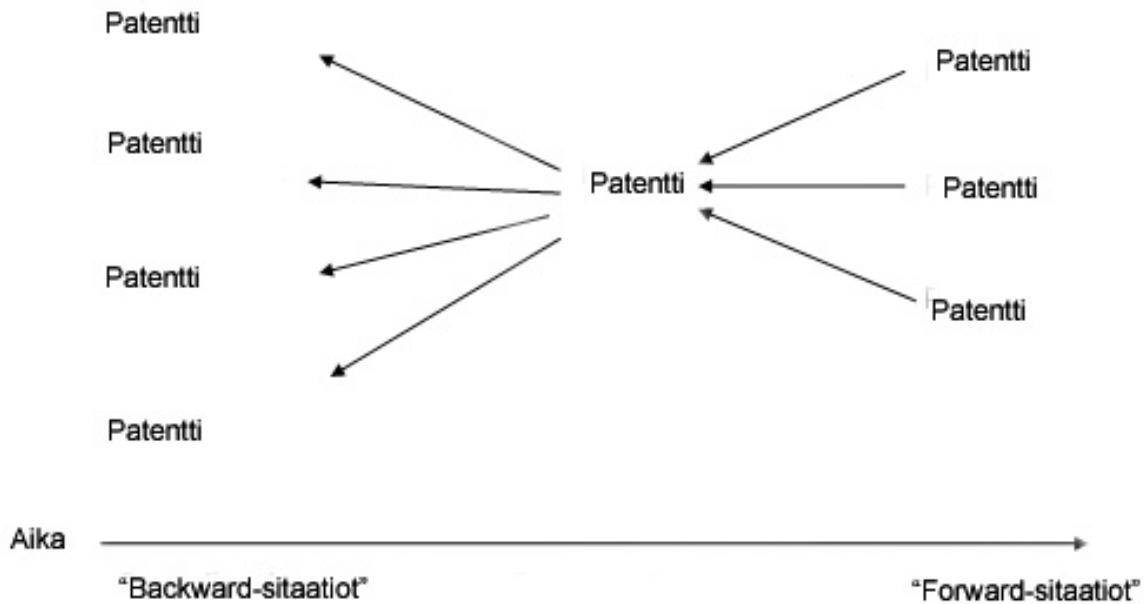
siten siis voida patentoida esimerkiksi löytöä, tieteellistä teoriaa, matemaattista menetelmää, taiteellista luomusta, suunnitelmaa, sääntöä tai menetelmää, peliä tai liiketoimintaa tai tietokoneohjelmaa (pl. Algoritmit). Keksinnöksi ei myöskään katsota kirurgista menetelmää. [11]. Patenttidataa ei ole voitu aiemmin hyödyntää suurissa määrin, koska dataa on valtavasti. Nyt laskentatehon kasvettua räjähdysmäisesti tietokoneissa, voidaan koko datamääriä ajaa erillisellä palvelimella tai pöytäkoneella. [9].

Patenttisitaatiot rajaavat omistusoikeuden laajuutta joka on myönnetty patentille. Eli jos patentti B sitatoi patenttia A, eli A:lla on jo olemassa olevaa tietoa jonka pohjalle patentti B rakentuu ja johon B:llä ei ole oikeutta. Patentin hakijalla on velvollisuus paljastaa kaikki tietoa jo tehdystä tutkimuksesta tai keksinnöstä, mutta patenttien tutkija lopulta päättää sen, mitä patenteja patentti sitatoi. Tämä siksi koska patenttien tutkija on asiantuntija alallaan ja hänen pitäisi siksi pystyä tunnistamaan aikaisempi relevantti tieto jota patenttiin yritetään kätkeä. [9].

Sitaatioiden tutkiminen ei kumminkaan ole aivan niin yksinkertaista että laskemalla suoraan sitaatioiden määrä, voitaisiin päätellä patentin merkittävyys. Ongelmana on se, että voimme tarkastella vain niitä sitaatioita jotka on jo tehty tutkittavaan dataan, emmekä vielä voi tietää tulevista sitaatioista. Emme siis voi vielä tietää onko esimerkiksi vuonna 1990 tehty patentti, joka on saanut 5 sitaatiota, vähemmän merkittävä kuin patentti joka on tehty vuonna 1980 ja saanut 10 sitaatiota. [9, 12].

Patenttisitaatiot levittävät tietoa kahdesta suurimmasta innovatiivisuuden näkökulmasta. Ensimmäisenä ovat linkit keksintöjen, keksijöiden, valtuutettujen välillä ajan ja paikan suhteen. Patenttisitaatiot mahdollistavat erityisesti kvantitatiivisen, tarkan tutkimuksen tiedon leviämisestä maantieteellisesti ja institutionaalisesti vertaillen. [12]. Esimerkiksi lähteessä [16] on tutkittu tiedon siirtymistä teknologisesti kehittyneistä maista kehitysvaiheessa oleviin maihin.

Työn loppuosassa teemme esimerkkihaun jonkin yrityksen pohjalta, jossa käsittelemme samalla kahta eri sitaatiotapausta. Ensimmäisessä tapauksessa tutkimme mitä patenteja haun tulokset sitatoivat ja toisessa tapauksessa tutkimme mitkä patentit sitatoivat hakumme tulosta. Ensimmäistä tapausta kutsutaan "backward-sitaatioksi" ja jälkimmäistä "forward-sitaatioksi".



Kuva1. Sitaatiot [23]

Yllä olevasta kuvasta käy ilmi sitaatiotyypit. Forward-sitaatiot, joita patentti saa, tarkoittavat patentin informaation olleen perustana jollekin tulevalle keksinnölle. Backward-sitaatiot jotka lähtevät jostain tietystä patentista, toimivat tärkeinä laillisina toimintoina, koska ne rajoittavat patentille myönnettyjä omistusoikeuksia. Jos siis patentti B sitoi patentti A:ta, A ilmentää tietoa jonka pohjalle patentti B pohjautuu ja jota patentti B ei voi vaatia itselleen. Muiden patenttien tekemien sitaatioiden (forward sitaatiot) voidaan ajatella kertovan patentin teknologisesta merkittävydestä. [13].

Eräs sitaatiotyyppi on yrityksen omiin patenteihin viittaaminen. Mikäli yritys viittaa lähinnä omiin hakemuksiinsa, voi yrityksen kehittää omaa tekniikkaansa tai se haluaa rakentaa muuria patenttiansa ympärille. [17].

2.2 IPC-luokitus

International Patent Classification (IPC) on hierarkkinen järjestelmä joka jakaa patentit yli 70000 eri kategoriaan. IPC-järjestelmä on WIPO:n ylläpitämä. Patenttiluokitus voi olla esimerkiksi "G02C 5/20". Seuraavina olevissa taulukoissa selvennetään IPC-luokitusta sekä sen alaluokkia.

Luokka	Kuvaus
A	Ihmisten tarvikkeet
B	Suoritettavat operaatiot, kuljetus
C	Kemia, metallurgia
D	Tekstiilit, paperi
E	Kiinteät rakenteet
F	Koneenrakennus, valaistus, lämmitys, aseet, räjäytys
G	Fysiikka
H	Sähkö

Taulukko 1. IPC-luokituksen pääluokat [18]

Tästä voidaan tarkastella seuraavaksi esimerkiksi luokkaa D, eli tekstiilejä sekä paperia. Alla olevasta taulukosta käy ilmi kyseisen luokituksen aliluokat.

Aliluokka	Kuvaus
D01	Luonnolliset tai keinotekoiset säikeet ja kuidut, kehrääminen
D02	Rihmat, rihmojen ja köysien mekaaninen viimeistely, loimet tai loimitukit
D03	Kutominen
D04	Palmikointi, nauhanteko, neulonta, trimmaukset, ei-kudottavat kuidut
D05	Ompelu, koruompelu
D06	Tekstiilien käsittely, pyykinpesu
D07	Köydet, kaapelit, muut kuin sähköiset
D21	Paperinvalmistus, selluloosan valmistus

Taulukko 2. D-pääluokan aliluokat [18]

Näin IPC-luokat jakautuvat yhä pienempiin ja pienempiin luokkiin, jolloin niillä saadaan luokiteltua patentit erittäin tarkasti.

2.3 Patentin käyttö

Patentilla on kaksi tärkeää käyttötarkoitusta, joista ensimmäinen on **suojaaminen**. Patentti antaa sen omistajalleen mahdollisuuden kieltää keksintönsä käytön kaupallisesti tietyssä maassa tai tietyllä alueella jonkin määrätyn ajan kuluksi. Yleensä tämä aika ei ole pidempi kuin 20 vuotta. Toinen käyttötarkoitus on **Tiedon jakaminen**. Patenttia päästään tutkimaan julkisesti, jotta se hyödyttäisi uusien teknologioiden kehittämistä ja taloudellista kasvua. [7].

2.3.1 Patenttihakemuksen täyttäminen

Patenttihakemuksen voi täyttää kolmella eri tavalla. Patentin voi siis täyttää **1) Kansallisesti** jolloin hakemus patentille täytetään yleensä kansallisessa patenttitoimistossa ja patentti voidaan myöntää vain maahan jossa sitä on haettu. **2) Alueellisesti**. Joillakin alueilla alueellisia patenttihakemuksia voi hakea alueellisista patenttitoimistoista, kuten African Regional Intellectual Property Organizationista (ARIPO) tai European Patent Officesta (EPO). Alueellisilla patenttihakemuksilla on sama vaikutus kuin alueen jäsenmaassa haetulla hakemuksella. **3) Kansainvälisesti**. Kansainvälisiä patenttihakemuksia voi täyttää toimistoissa jotka kuuluvat Contracting States of the Patent Cooperation Treaty (PCT) -organisaatioon tai WIPO:n kansainvälisessä toimistossa (International Bureau of the World Intellectual Property Organization). Hakemuksen voi täyttää kuka tahansa asukas tai kansalainen joka asuu PCT:n vaikutusalueella. Yhdellä kansainvälisellä patenttihakemuksella on sama vaikutus kuin kansallisilla hakemuksilla, jotka jokainen täytettäisiin halutuissa PCT-valtiossa. [7]

Patentti on siis tarpeen niissä maissa, joissa keksintöä aiotaan hyödyntää ja niissä, joissa on odotettavissa kilpailevaa valmistusta. [11].

2.4 Patentin hakuprosessi

Patentin hakuprosessi vaihtelee eri toimistojen kesken, mutta pääsääntöisesti prosessi on seuraavanlainen:

Täyttö. Hakija määrittää hakutyypiksi kansallisen, kansainvälisen tai alueellisen ja täyttää sitten hakemuksen. Ensimmäinen haku on tyypiltään ”priority filing” josta jokin edellä mainituista hakutyypeistä valitaan ”priority periodin” aikana joka on yksi vuosi. [7].

Muodollinen tutkinta. Patenttitoimisto varmistaa että kaikki muodollisuudet, kuten tarvittavat dokumentit löytyvät, täyttyvät ja maksut on maksettu. [7].

Aikaisempiin tutkimuksiin tutustuminen. Lähes kaikkien maiden patenttitoimistoissa tehdään tutkimus edeltävästä tiedosta liittyen haettavaan patenttiin. Eli etsitään kaikki patenttiin liittyvä tieto. Laajojen tietokantojen ja asiantuntijoiden avulla tehdään tutkimusraportti joka vertaa patentin väitettyä keksintöä jo olemassa oleviin. [7].

Julkaisu. Suurimmissa osissa maita patenttihakemus julkaistaan 18 kuukautta sen ensimmäisestä hakupäivämäärästä. [7].

Patentoitavuuden vaatimukset

Jos aikaisempaa tietoa on saatavilla, tutkija tarkistaa että hakemus täyttää patentoitavuuden vaatimukset, eli keksintö on uudenlainen sekä kekseliäs teollisuuden hakemus, verrattuna aikaisempien tutkimusten tuloksiin, jotka on listattu tutkimusraportissa. Tutkija voi myöntää patentin ilman muutoksia, muuttaa sen vaatimuksen laajuutta siten että se vastaa/heijastaa jo tunnettuun tutkimukseen tai hän voi hylätä hakemuksen. [7].

Vastustaminen. Monet patenttitoimistot sallivat määrätyn ajan sisällä tapahtuvan vastustuksen patenttia kohtaa, mikäli se ei täytä sille asetettuja vaatimuksia. Kolmannet osapuolet voivat vastustaa patentin myöntämistä. [7].

Päätöksestä valittaminen. Jos patenttia vastustetaan, monet patenttitoimistot sallivat myös patentin myöntämisestä johtuvan päätöksestä valittamisen. [7].

2.5 Miksi patenti-informaatiota

Patenttitieto on tärkeää tutkijoille, keksijöille, viihdyttäjille ja kaupallisille yhtiöille sekä patenttialan ammattilaisille. Patenttitietojen avulla vältetään tekemästä turhaa tutkimusta ja kehitystä koska saadaan tietoa jo olemassa olevista keksinnöistä sekä vältetään rikkomasta muiden keksintöjen tietoja. Saadaan myös tietoa muiden keksijöiden tekemistä merkittävistä patenteista sekä sellaisista patenteista, joita ei koskaan ole myönnetty, eivät ole voimassa jossain tietyssä maassa tai niiden voimassaoloaika on jo päättynyt. Voidaan myös seurata teknologian kehittymisen trendejä ja suuntauksia, kuten sellaisia jotka liittyvät terveydenhuoltoon tai ympäristöön ja tuottaa perustaa toimintatapojen suunnitteluun [7].

2.5.1 Mitä informaatiota patenttidokumentti sisältää?

Patenttidokumentti sisältää teknistä tietoa, joka kuvailee keksintöä ja siihen voi liittyä myös piirustuksia. Juridista tietoa jolla määritetään patentin laajuus ja sen laillinen status. Business-tietoa, kuten keksijän nimi, hakupäivämäärä ja maa jossa hakemus on tehty. Julkisiin menettelytapoihin liittyvää tietoa joka saadaan analysoimalla patenttien täyttöä, ja jota päättäjät voivat käyttää hyväkseen kansallisissa teollisuuden menettelystrategioissa. [7].

2.5.2 Mistä patenttitietoa voi löytää?

Patenttitietoja on julkaistu muutamien julkisten tietokantojen avulla. Jokainen tietokanta kattaa tietyn otannan tietoja, eikä tällä hetkellä yhdessäkään tietokannassa ole kaikkea maailman patenttitietoa, joten mahdollisimman tarkan tutkimuksen kannalta pitäisi tutkia monta tietokantaa [7].

On olemassa ns. Offline- ja online-tietokantoja. Online-tietokannat ovat Internet-sivustoja, joiden kautta voi tehdä erilaisia hakuja eri patenttitietokannoista. Tällaisia sivustoja ovat mm. Espacenet (<http://fi.espacenet.com>), US-patent (<http://www.uspto.gov>) ja freepatensonline (<http://www.freepatensonline.com>). Espacenetin sivustolta voi hakea patenteja hakusanoilla,

julkaisunumeroilla, päivämäärällä, hakijalla ja keksijällä. US-patentin sivustolta voi hakea USA:n patenttien täydet tekstit esimerkiksi jollain tietyllä hakusanalla. Freepatensonline sisältää voimassaolevat US-patentit pdf-tiedostoina.

Myös monet kansalliset ja alueelliset patenttitoimistot tarjoavat patentti-informaatiota Internetin välityksellä. Lähteessä [22] on lueteltu tarkemmin eri maiden toimistot ja mainittu myös patenttitoimistojen hakukoneet. Hakukoneet käsittävät tietoa vain kunkin toimiston alueelta. [7, 22].

Offline-tietokannat eroavat suuresti online-tietokannoista. Ne ovat yleensä tietokantoja, joita varten joudutaan pystyttämään oma palvelin josta tietoa saadaan haettua, mikäli koko datamäärää ei ajeta käyttäjän koneelta. Etu näissä tietokannoissa verrattuna online-tietokantoihin on se, että tietoa saadaan suuressa määrin analysoitavaksi. Tietokantojen avulla voidaan hakea esimerkiksi jonkin tietyn yrityksen kaikki patentit ja tehdä niistä tarkempaa tutkimusta. Näitä käytettäessä ei siis perehdytä yksittäisiin patentteihin vaan suurempaan määrään tietoa. Tietenkin tietokannan tiedoista voi saada hieman tarkempaakin tietoa kuten abstract-kentän, mutta harvemmin kokonaisia patenttitietoja. Huomioitavaa offline-tietokannoissa on myös niiden hinta. Itse tietokannan tiedot joudutaan ostamaan ja päättämään kuinka palvelimen tekninen toteutus suoritetaan vai ostetaanko se palveluna.

3 PATSTAT-tietokanta

Tässä kappaleessa esittelemme Patstat-tietokantaa ja käymme läpi asioita ja ohjelmia joita otimme huomioon tietokannan teknisessä toteutuksessa käytettävyyden ja suorituskyvyn kannalta. Esittelemme myös muutaman ongelman tietokannan teknisessä toteutuksessa.

Patstat esiteltiin ensimmäisen kerran huhtikuussa 2006. Tietokantaa päivitetään kaksi kertaa vuodessa ja sen jakelumuotona on DVD-levyt. Levyt sisältävät jaettavat tiedot CSV-formaatissa ja tietojen koko on noin 100 gigatavua indekseineen. Tietokanta on toteutettu relaatiomallisella rakenteella ja se sisältää 15 taulua . [5].

EPO on Patent Statistics Task Forcen aktiivinen jäsen jota OECD johtaa. Jäseniä ovat:

- ECD, Organisation for Economic Co-operation and Development (chair)
- EPO, European Patent Office
- JPO, Japan Patent Office
- USPTO, Unites States Patent and Trademark Office
- WIPO, World Intellectual Property Organisation
- NSF, National Science Foundation (US)
- Eurostat (European Commission)
- DG Research (European Commission)

Työryhmä on pyytännyt, että EPO tekisi maailmanlaajuisen patenttitietokannan. [6].

3.1 Tekninen toteutus

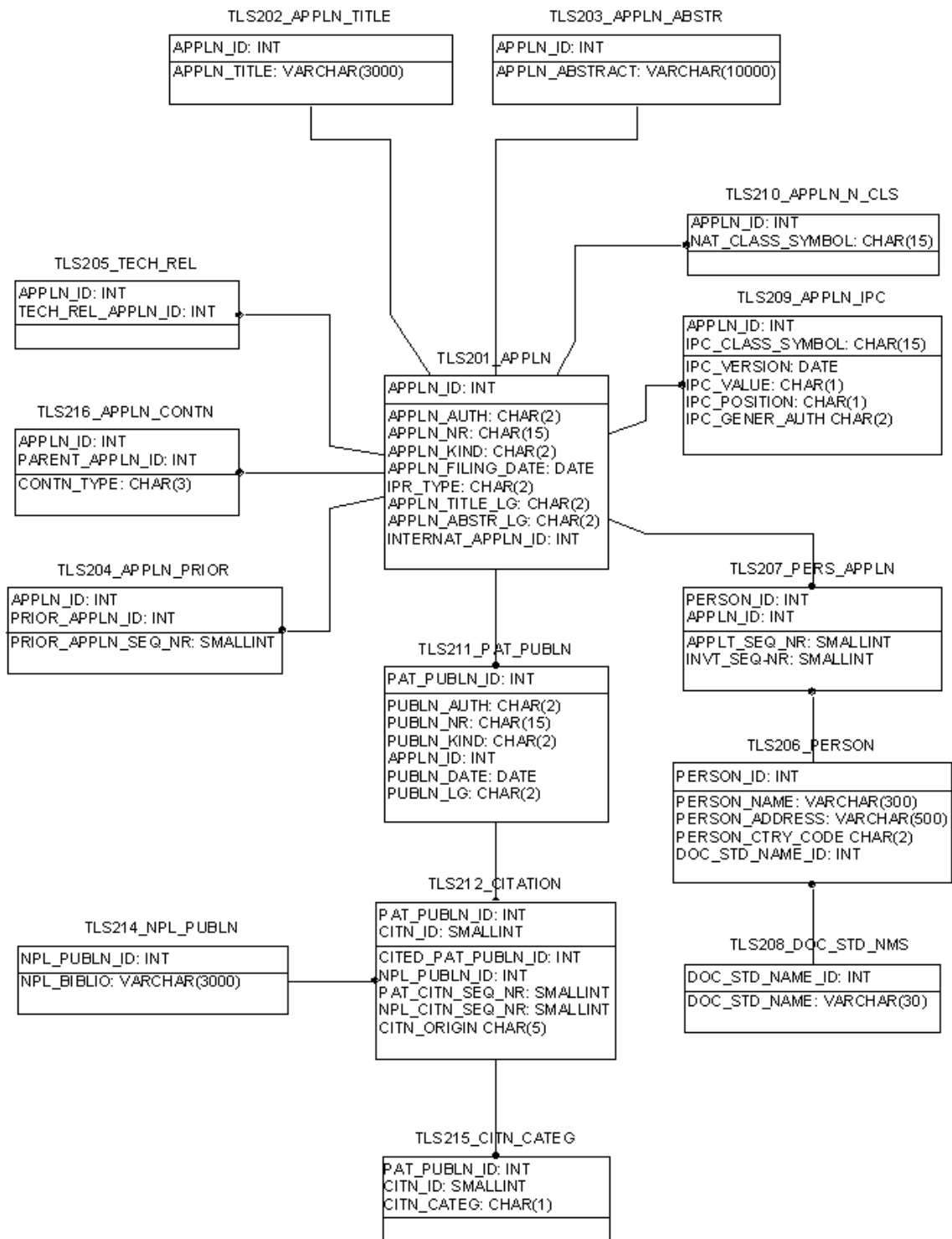
Tietokantaa ei päätetty ostaa palveluna vaan toteuttaa se itse hankkimalla tarvittava palvelin, asentamalla siihen käyttöjärjestelmä ja tarvittavat ohjelmistot tietokannan toiminnan kannalta. Palvelimen hankintaan päädyttiin toteamalla sen olevan edullisempaa, kuin sen ostaminen palveluna, jo parin vuoden käytön jälkeen. Tutkimuslaboratorio Chair of Economics and Mangement of Innovation (CEMI, 2009) on käyttänyt omassa Patstat-tietokantaserverissään

seuraavanlaista kokoonpanoa: Intel Xeon prosessori 2 GHz kellotaajuudella, 3 gigatavua RAM-muistia sekä 2 kappaletta SCSI-liitäntäisiä kovalevyjä joiden pyörimisnopeus on 15000 rpm [8].

Palvelimen toteutuksessa päädyimme seuraavanlaiseen kokoonpanoon: Intel Xeon prosessori, jossa 4 ydintä 2 GHz kellotaajuudella, 8 gigatavua RAM-muistia ja 2 kappaletta SAS-kovalevyjä 15000 rpm kierrosnopeudella. Palvelin on ns. ”räkkipalvelin” eli telineeseen asennettava palveli, jota voi tulevaisuudessa laajentaa toisella prosessorilla ja lisäämällä muistia tai kovalevyjä. Palvelimen on tässä tapauksessa hyvä olla riittävän tehokas tarpeellisen suorituskyvyn saavuttamiseksi sillä useampi tutkija voi käyttää palvelinta samanaikaisesti suorittaen paljon laskentatehoa vaativia kyselyitä. Omien kokemuksiemme mukaan tietokannan suurimmat taulut ovat useiden kymmenien miljoonien rivien suuruisia, joka myös asettaa laitteistolle kovemmat vaatimukset.

Tietokannan tietoturvaa emme siis käsittele tässä työssä, mutta mainitsemisen arvoista on se, että tietokanta toimii yliopiston sisäisessä verkossa ja siihen saa yhteyden ulkopuolelta vain VPN-yhteyden ylitse. Taulujen sisältämät tiedot ovat julkista tietoa joten niiden suojaamiseen ei ole hyödyllistä kuluttaa aikaa tai muita resursseja.

Seuraavalla sivulla olevasta kuvasta käy ilmi tietokannan taulujen rakenne ja liitokset.



Kuva 2. Patstat-patenttitietokannan relaatiomallin rakenne [8].

3.1.1 phpMyAdmin

Palvelimeen päätettiin asentaa phpMyAdmin -niminen ohjelma, jonka avulla pystytään esimerkiksi hallitsemaan tietokantaa sekä lisäämään käyttäjiä. Ohjelmassa on yksinkertainen käyttöliittymä, jonka avulla on tarkoitus myös suorittaa tietojen hakeminen SQL-kyselyillä. Harkinnassa oli ensin toteuttaa erillinen käyttöliittymä php:n avulla, mutta koska tutkijat haluavat tehdä niin monia erilaisia kyselyitä, on niiden syöttäminen SQL-muodossa helpoin toteutus. PhpMyAdmin myös sallii tietojen viemisen monissa eri muodoissa jatkokäsittelyä varten. Tärkeimmäksi tiedostomuodoksi viemisen kannalta muodostui CSV, sillä se on helpohko siirtää Microsoft Accessiin, joka löytyy käytettävistä työasemista.

3.1.2 RAID

RAID (Redundant Array of Independent Disks) on tekniikka jolla useita tietokoneen kovalevyjä voidaan yhdistää yhdeksi eri tavoin. Kaksi yleisimmiten käytettyä RAID-kokoonpanoa ovat RAID 0 ja RAID 1. RAID 0 tarkoittaa kahden levyn yhdistämistä yhdeksi loogiseksi levyksi. Tällainen kokoonpano on nopea ja tuottaa suurimman nopeuden lisäyksen järjestelmään mm. nopeuttamalla haku-aikaa. Haittapuolena on se, että jos toinen käytettävissä olevista levyistä hajoaa, menetetään kummankin kovalevyn sisältämä data. RAID 1 tarkoittaa peilaamista jolloin kummallakin kovalevyllä säilytetään sama data ja toisen rikkoontuessa ei dataa menetetä ollenkaan. Patstat-käyttöön valitsimme RAID 0 -kokoonpanon suurimman nopeuden saamiseksi. [1].

3.1.3 Tietokantamoottoreista

Kun luodaan MySQL-pohjaisia palveluita tai ohjelmistoja, täytyy valita käytettävä tietokantamoottori tietokannalle. MySQL:n tukemia tietokantamoottoreita ovat mm. InnoDB, BDB ja MyISAM [1]. Eri moottorit soveltuvat toisistaan poikkeaviin käyttötarkoituksiin. Esimerkiksi InnoDB soveltuu käyttöön, jossa tarvitaan transaktioita sekä luku- ja kirjoitustapahtumien yhtäaikaaisuutta. MyISAM taas soveltuu käyttöön joka ei vaadi transaktioita vaan lähinnä SELECT-

ja INSERT/UPDATE -kyselyitä. [1, 2] Tämän vuoksi tietokannan teknisessä toteutuksessa otettiin MyISAM -tietokantamoottori käyttöön, sillä tietokantaan tehdään lähes pelkästään SELECT-kyselyjä tietoja haettaessa ja nopea tulosten ulos saaminen onkin hyödyllistä kun kyseessä on näin suuri tietokanta. Patstat-tietokantaa ei tarvitse päivittää kovin usein, eikä sinne ole tarvetta lisätä tietoja. Mikäli patenttitietoja on syytä ajanmukaistaa, voi päivityksille varata riittävästi aikaa sillä tietokannan ei tarvitse olla käytettävissä 24/7.

3.1.4 Indeksointi

Tietokannan taulujen indeksejä voisi verrata puhelinluettelon käyttöön. Jos haluaisimme etsiä puhelinluettelosta jonkin tietyn nimen, emme aloittaisi selaamaan luetteloa alusta alkaen vaan hyppäisimme jonkin tietyn kirjaimen kohdalle. Indeksit tietokannoissa toimivat samoin tavoin. Mikäli haluaisimme etsiä käyttäjän "Pekka" tietoja taulusta, indeksit ohjaavat tietokantamoottorin aloittamaan taulun selaamisen P-kirjaimen kohdalta. Ilman indeksejä tietoa lähdetäisiin etsimään taulun alusta, eikä se olisi tehokasta ja suurissa tietokannoissa hidastaisi toimintaa erittäin paljon. [1]. Koska Patstat-tietokannassa tiedot on jaettu yli 10 eri tauluun, on tärkeää että taulujen liitoksina toimivat tiedot on indeksoitu oikein suorituskyvyn varmistamiseksi.

Indeksoinnin huono puoli on se, että tietojen lisääminen tauluihin, joissa on jo tietoa, on olemassaolevien indeksien takia hidasta. Patstat-tietokannan tapauksessa tällä ei tosin ole mitään merkitystä, sillä tietoa on lisättävänä harvoin eli pelkästään päivitysten yhteydessä ja sille voidaan antaa hyvin aikaa tekemällä päivitykset esimerkiksi viikonloppuisin, jolloin tutkijat tai jatko-opiskelijat eivät todennäköisesti tarvitse patenttitietoa. Hitaus johtuu siitä, että indeksit ovat eri sarakkeissa kuin itse tiedot (indeksit vain viittaavat tietoihin) ja kun tietoja päivitetään, on myös itse indeksit päivitettävä [1].

3.1.5 Esille nousseet ongelmat Patstat-tietokannassa

Patstat-tietokannan käytössä on ilmennyt joitakin puutteita tai heikosti toteutettuja osa-alueita. Ensimmäisenä mainittakoon ns. "nimiongelmat". Jokaiseen patenttitietoon kuuluu siis yrityksen tai henkilön nimi, jolle patenttia on haettu. Yrityksen nimet saattavat joissakin tapauksissa muuttua

useinkin niiden historian aikana, esimerkiksi UPM Kymmene ja Raflatac viittaavat Patstatin kohdalla samoihin tietoihin. Koska itse tietokannan tiedoissa ei ole mitään mainintaa tai viittausta yrityksen nimihistoriaan, on tietojen käsittelijän se itse selvitettävä. Yritys saattaa esiintyä myös muutamalla samankaltaisella nimellä. Esimerkiksi UPM Kymmene esiintyy seuraavilla nimillä: UPM Kymmene, UPM Kymmene Corp, UPM Kymmene Oy, UPM Kymmene Oy Helsinki, UPM Kymmene Oy Valkeakoski, UPM Kymmene Walki Pack, UPM Kymmene Oyj, UPM Kymmene Oyj Helsinki, UPM Kymmene Oyj Kaukas Kemiall, UPM Kymmene Oyj Lappeenranta, UPM Kymmene Oyj UPM Kymmene Oy, UPM Kymmene Papier GMBH & CO K, UPM Kymmene Wood AB, UPM Kymmene Wood Oy ja UPM Kymmene Wood Oy Lahti.

Usein tietoja haetaan tietokannasta ns. "jokerimerkkien" avulla, eli haetaan yrityksen nimen sisältävän taulun tietystä sarakkeesta tiettyä sanaa. UPM:n tapauksessa jokerimerkkiä on helppo käyttää kun se sijoitetaan tekstin "Upm kymmene" loppuun: "upm kymmene%". Tämä toisaalta vaatii yrityksen nimen tarkistamisen etukäteen tietokannasta. Toinen hyvä vertailukohde voisi olla yritys nimeltä "Tego". Tässä ilmeni ongelma siten, että haettaessa SQL-kyselyllä kyseistä tekstinpätkää kaikista kohdista merkkijonoa, tuloksessa esiintyi yli puolet epärelevantteja tuloksia. Tego tapauksessa tuloksiin tuli erittäin paljon Japanilaisia patenteja. Tämä aiheuttaa täten tietyissä tapauksissa paljon työtä datan käsittelijältä, väärrien tulosten poissaamiseksi.

Patenttitietokannassa myös IPC-luokkien toteutuksessa on toivomisen varaa. IPC-luokat ovat siis hyvin tarkkoja merkintöjä patenttien luokitukselta, esim: " H01L 21/762". Mikäli haluaisimme muodostaa haun, jonka ehtona olisi edellä mainittu luokitus, on huomattava että tekstijonossa esiintyy kaksi tyhjää merkkiä L-kirjaimen ja ensimmäisen numeron välissä. Jos otetaan toiseksi esimerkiksi luokitus " G05F 1/10", huomataan että samassa välissä on tällä kertaa kolme tyhjää merkkiä. Toteutus ei välttämättä ole huonoin mahdollinen, mutta se voi aiheuttaa SQL-kyselyjä suorittaville tutkijoille ongelmia, mikäli he eivät ole perehtyneet tietokannan dokumentaatioon.

Viimeinen mainitsemisen arvoinen esille noussut ongelma liittyy patenttitietojen lataamiseen tietokantaan palvelimen asennusvaiheessa. Asennuslevyinä toimivien DVD-levyjen mukana toimitettiin myös yksinkertaiset ohjeet joiden avulla patenttitietokanta saadaan pystytettyä ja toimimaan. Valitettavasti nämä ohjeet käsittivät asennuksen vain Microsoft Windows -käyttöjärjestelmiin, eikä mainintaa muista, esimerkiksi Linux-käyttöjärjestelmistä löytynyt.

Kokeneemmalle asentajalle tämä ei tuota suuria ongelmia ja tämänkin työn tapauksessa suurin työ oli tietojen lataamiseen käytettyjen SQL-skriptien sisältämien tiedostopolkujen muuttaminen Windows-muodosta Linux-muotoon.

4 TIETOKANNAN HYÖDYNTÄMINEN TUTKIMUSKÄYTÖSSÄ

Tässä kappaleessa käymme läpi kahdella eri tavalla toteutetut case-esimerkit tietokannan hyödyntämisestä. Ensimmäisessä kohdassa teemme haun kahden yrityksen nimen perusteella, eli haemme esimerkiksi kaikki Nokian patentit joista teemme analyysia. Toisessa analyysissa haemme patenteja pelkkien patenttiluokitusten, eli IPC-luokitusten avulla, jolla voimme tutkia mm. tietyllä teknologian osa-alueella toimivien yritysten patenteja. Havaitsimme myös, että tietokannan hyödyntäminen vaatii jonkinlaista SQL-kielen tuntemusta koska tutkijoilla voi olla hyvinkin erilaiset tarpeet hakujen suorittamiseen liittyen.

4.1 Yrityksen nimen perusteella

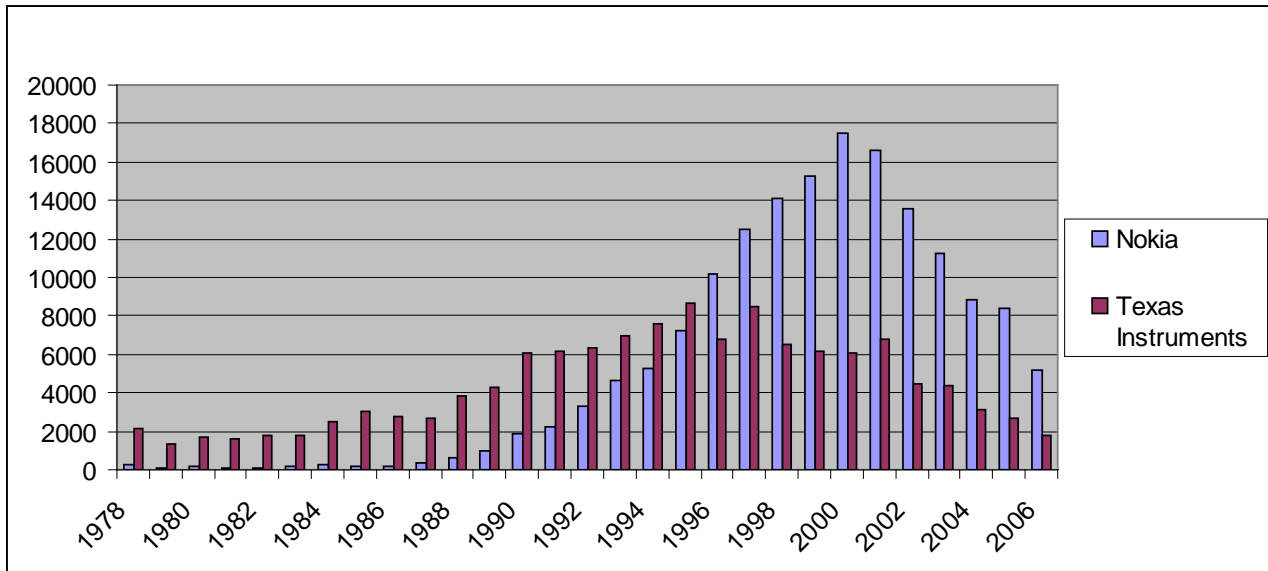
Yrityksen nimen perusteella tapahtuvassa haussa haetaan itse yrityksen nimen lisäksi seuraavat tiedot patenteista: patentin numero joka yksilöi sen tietokannassa, patenttihakemuksen täyttöpäivämäärä, IPC-luokitus, patentin julkaisunumero sekä mahdollisen viitattavan patentin numero. Näillä tiedoilla voidaan tutkia mihin teknologian osa-alueelle yritys on tehnyt patenteja tiettyinä vuosina ja kuinka patentoinnin osa-alueet ovat mahdollisesti muuttuneet.

Yritystietoja voidaan hakea myös useammista yrityksistä kerralla. Tämä helpottaa tutkijoiden työtä siten, että tietoja ei tarvitse yhdistellä eri hakujen jälkeen. Näin voidaan hakea esimerkiksi 20 yrityksen tiedot jotka toimivat paperiteollisuuden alalla.

Ongelmana useamman yrityksen hauissa on tiedon suuri määrä jatkokäsittelyä ajatellen. Tiettyjen hakujen kohdalla on huomattu tuloksena tulevan miljoonia rivejä tietoa. Sitaatiohaut lisättynä näihin tietoihin saavat jatkokäsittelyssä käytetyn Microsoft Accessin ääri rajoilleen sillä Accessin tietokantatiedoston maksimikoko on 2 gigatavua [10]. Muutamissa tapauksissa sitaatiohaut onkin jouduttu lisäämään erillisiin tiedostoihin. Tämä aiheuttaa tutkijoille lisää työtä.

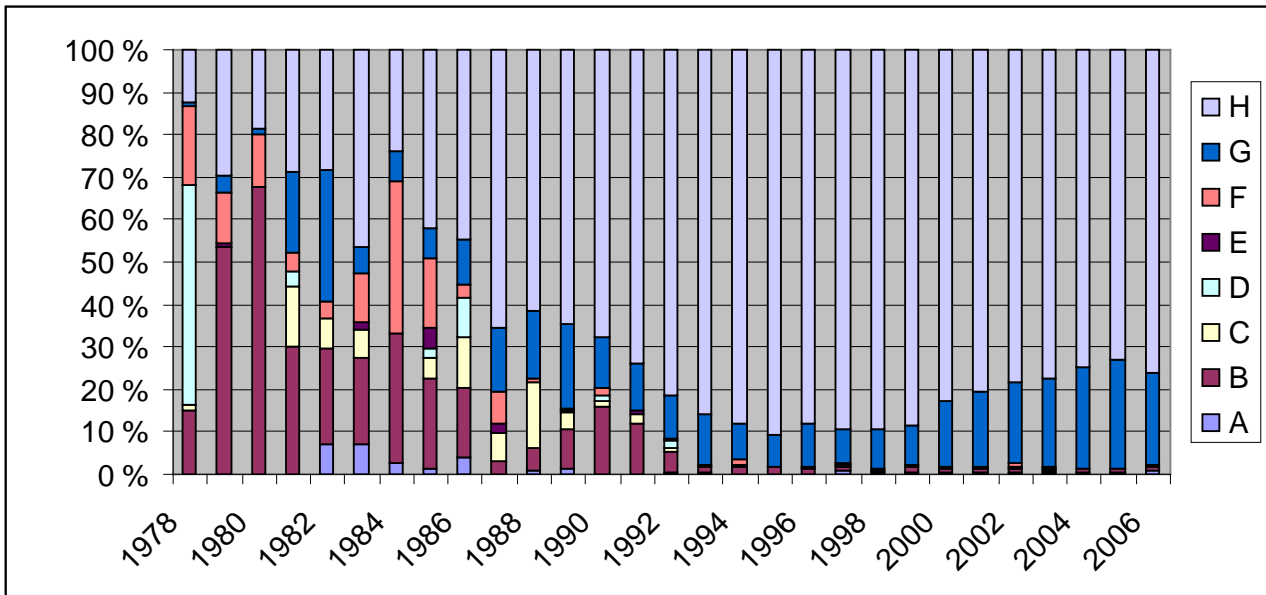
Seuraavassa esimerkkitapauksessa haimme kaikki Nokian sekä Texas Instrumentsin patentit. Saaduista tuloksista saimme aikaiseksi alla olevat kuvat, joista käy ilmi yritysten tekemien

patenttien määrä sekä IPC-luokitukset jaoteltuina vuosittain. IPC-luokkia käsitellään yleensä neljän ensimmäisen merkin kannalta, sillä se on riittävä tarkkuus analyysiä varten, esimerkkikuvat on otettu vain pääluokituksen eli ensimmäisen merkin kohdalta. Kuvien aineistot ovat vuosien 1978 ja 2006 väliseltä ajalta. Kuvasta 3 käy ilmi Nokian ja Texas Instrumentsin tekemien patenttien määrä vuosittain.

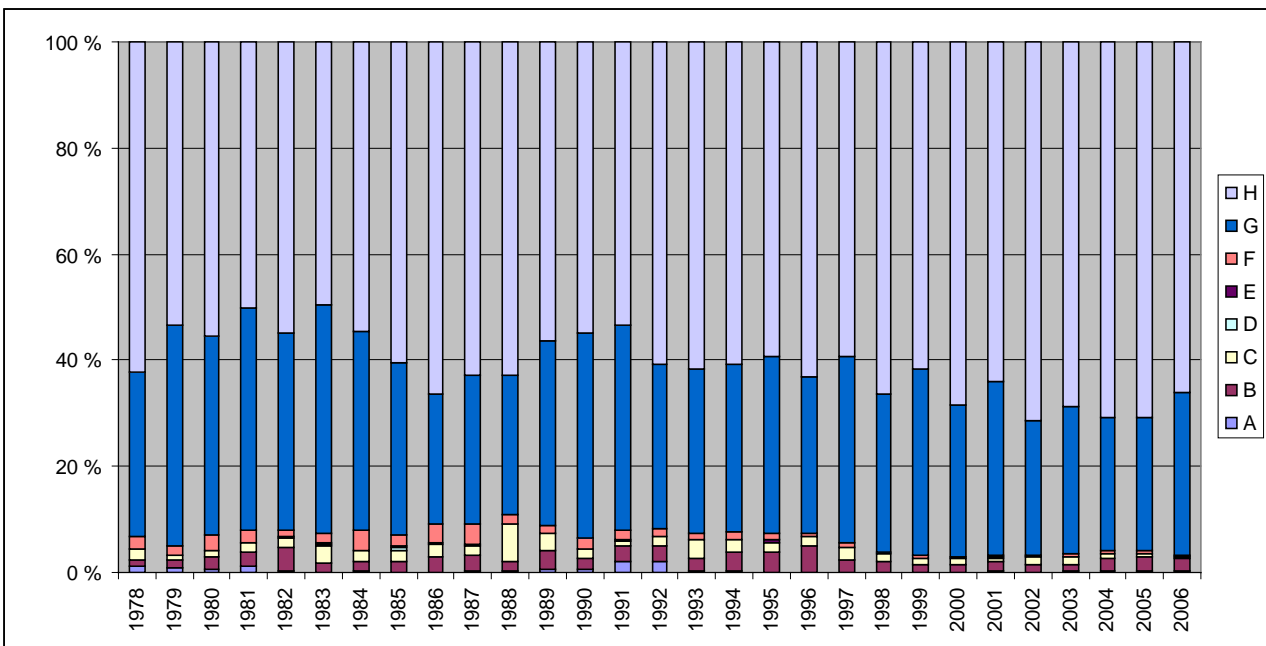


Kuva 3. Nokian sekä Texas Instrumentsin patenttimäärien vertailu

Kuvaa tutkiessa on myös muistettava, että patenteihin liittyy tietty viive joilla ne julkaistaan [9]. Kuvasta voisi saada käsityksen yrityksen patentoinnin hiipumisesta, mutta asia ei ole näin. Seuraavalla sivulla olevasta kuvasta 4 käy ilmi Nokian tekemien patenttien määrän suhteellinen vuosittainen jakauma IPC-luokkien perusteella. Kuvasta on hyvin havaittavissa yrityksen kehityssuunnan muutos 1980-luvun loppupuolelta alkaen. Kuvassa 5 esitetään sama jakauma Texas Instrumentsin patenteista.



Kuva 4. Nokian patenttien vuosittainen jakauma IPC-luokkien perusteella.

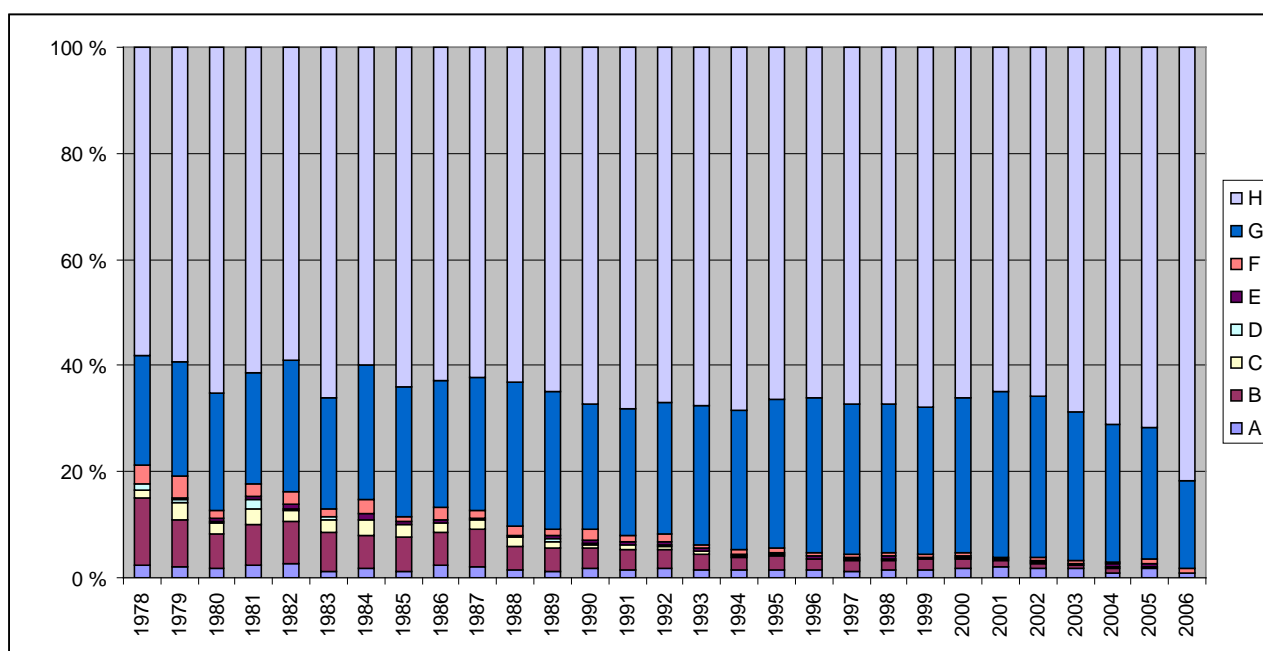


Kuva 5. Texas Instrumentsin patenttien vuosittainen jakauma IPC-luokkien perusteella.

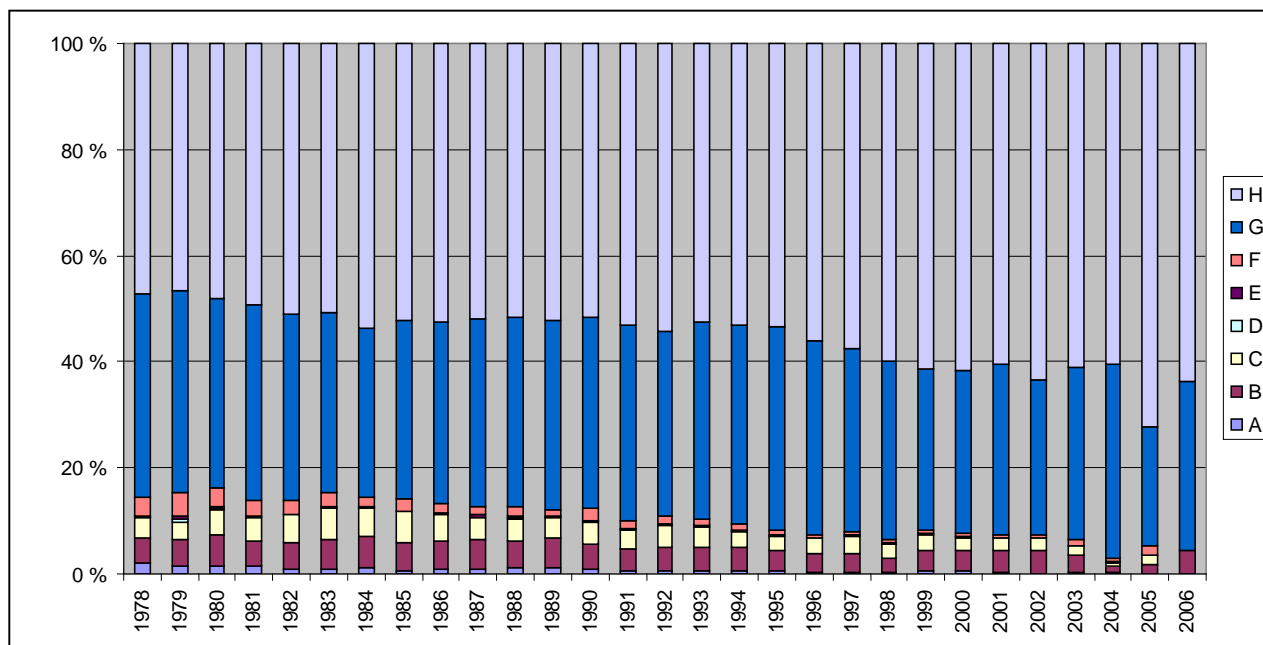
4.1.1 Sitaatioista

Kun haut on tehty yrityksen nimen perusteella, pitää tiedot viedä Microsoft Accessiin jatkokäsittelyä varten. Accessissa haetuista tiedoista erotellaan patentin julkaisunumero ja viitatun patentin numero omiin tauluihinsa ja viedään ne CSV-formaattiin. Nämä tiedostot viedään tietokantaan omiksi tauluikseen jota kautta rakennetaan haku ns. ”backward” ja ”forward” -sitaatioita varten. Sitaatiohaut voisi tehdä ilman erillisten taulujen rakentamistakin, mutta se olisi omien kokemustemme mukaan palvelimen suorituskyvyn kannalta huonompi toteutus. Erillisiin tauluihin viemällä tiedoista saadaan myös poistettua epärelevantit tulokset, kuten väärin yritysten nimet. Näin myös patenttitutkimuksen tarkkuus saadaan paremmaksi.

Saatujen tulosten perusteella voidaan analysoida mihin IPC-luokkiin tietyn yrityksen patentit viittaavat. Tämän perusteella voidaan myös tutkia, että viittaavatko yrityksen patentit oman toimialansa luokituksiin vai toisen toimialan luokituksiin. Seuraavaksi esitettävissä kuvissa 6 ja 7 on patenttitietokannasta saadut tulokset, joissa käsitellään IPC-luokituksia joihin Nokia sekä Texas Instruments ovat viitanneet. Sitaatioiden tulokset viittaavat patenteihin joiden hakupäivämäärä on aikavälillä 1978-2007.



Kuva 6. Nokian tekemät backwards sitaatiot.

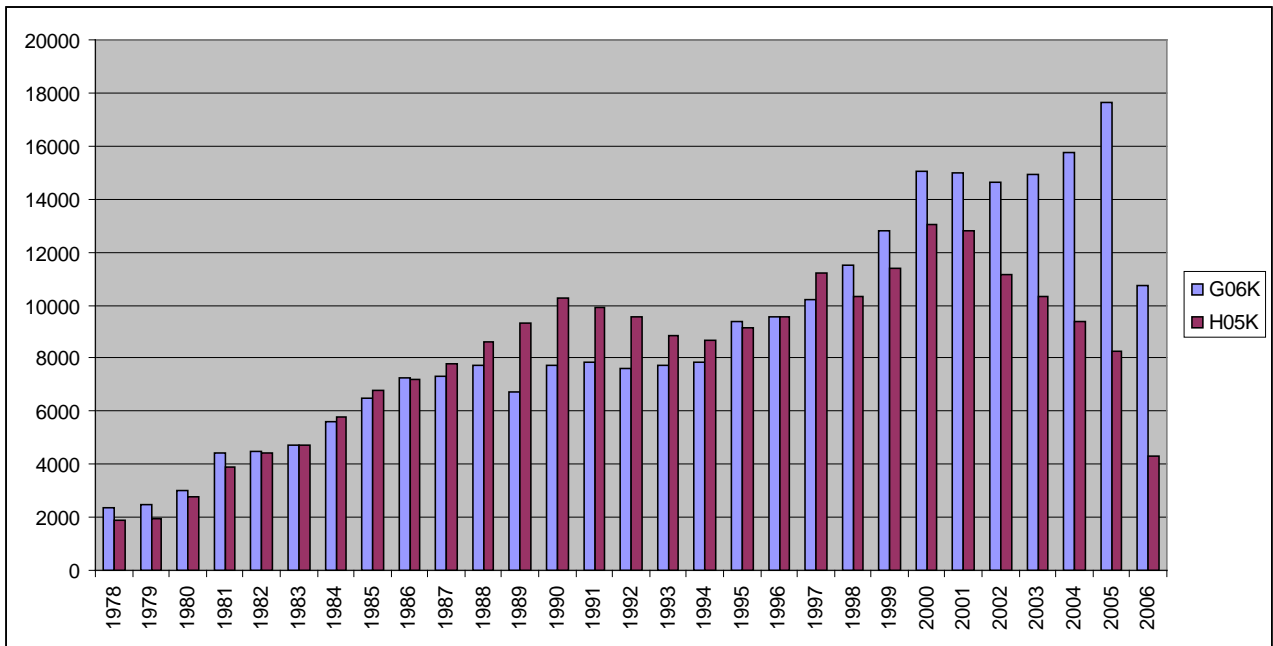


Kuva 7. Texas Instrumentsin tekemät backwards-sitaatiot.

Tähän mennessä on siis tehty ns. ”perushaku” josta löytyy viitattavan patentin numero ja sitaatiohaku, jonka tuloksena on patentit joihin ensimmäinen haku viittaa. Nämä tiedot voidaan yhdistää Accessissa jolloin nähdään mikä patentti viittaa mihinkin. Näin tarkkaa tutkimusta ei kumminkaan ole tarkoitus tehdä vaan painopiste on tutkia millaisiin patenteihin tietyt yritykset tai yritysryhmät viittaavat.

4.2 Patenttiluokitusten avulla

Patenttiluokitusten, eli IPC-hakujen avulla saadaan kerralla tietoa kaikista yrityksistä, jotka ovat patentoineet tietylle teknologian osa-alueelle. Tietoa voidaan käyttää esimerkiksi vertailemaan eri yritysten patenttien määriä keskenään tai vertailemaan eri IPC-luokitusten patenttimääriä vuosittain. Seuraavalla sivulla olevasta kuvasta selviää H05K (piirilevyt; kuoret tai rakenteelliset yksityiskohdat) ja G06K (Datan tunnistus; datan esitys) luokkien vuosittainen patenttijakauma. Luokitukset valittiin tarkasteltavaksi meneillään olevan tutkimuksen takia, johon liittyy monia RFID -toimijoita, kuten edellä käsitellyt Nokia ja Texas Instruments.



Kuva 8. H05K ja G06K patenttiluokitusten jakautuminen vuosittain.

5 JOHTOPÄÄTÖKSET

Tässä luvussa esitetään havainnot tutkimuskysymykseen valituista näkökulmista. Tutkimusongelmana oli patenttitietokannan tietojen hyödyntäminen tutkimuskäytössä. Tutkimus kohdistui patenttitietojen rakenteen selvittämiseen, patenttitietokantoihin tutustumiseen sekä tietojen hyödyntämisen selvittämiseen.

Teoriaosuudessa käsitelimme sitaatiohakuja ja niiden merkitystä tutkimuksen kannalta. Tietokannan avulla lähdimme etsimään ratkaisuja näihin asioihin ja voimme selvittää mm. patenttien sitaatiot monella eri tavalla. Sitaatioita voidaan käsitellä eteen- tai taaksepäin suuntautuneina ja "self-sitaatteina". Pystymme myös tarkastelemaan viitattujen patenttien luokituksia jolloin voimme selvittää viittaavatko patentit toimialansa sisäisiin vai ulkoisiin patentteihin.

Muutaman kerran ongelmaksi nousi tiedon jatkokäsittelyyn liittyvät ongelmat, kuten Microsoft Accessin rajoitteet ja tietokannan taulujen tiedot joiden toteutus ei ollut paras mahdollinen. Suuren tietomäärän takia kaikkia tietoja ei saanut jokaisessa tapauksessa mahdutettua yhteen Accessin tietokantatiedostoon, joka aiheutti lisätyötä.

Tietokannan toteutuksesta johtuen tutkijat joutuvat tietyissä tapauksissa käsittelemään tietoja jonkin verran manuaalisesti esimerkiksi poistamalla vääriä hakutuloksia ja viemällä tietoja tauluihin erillisiin tietokantoihin. Tämä vaatii tutkijoilta jonkin verran teknistä tietämystä jatkokäsittelyssä käytettävien ohjelmistojen suhteen, mutta tällä varmistetaan tutkimustulosten tarkkuus eivätkä epärelevantit tulokset sotke niitä. Tietokannan käyttö vaatii hieman SQL-kielen tuntemusta ja opettelua, mutta jo muutama eri haku yhdistettynä Accessin käyttöön tuottaa erittäin paljon dataa tutkimusta ajatellen.

Patenttien sitaatioiden lisäksi tietokannan avulla voidaan helposti tutkia haluttujen yritysten tekemiä patenttimääriä sekä patenttien luokituksia. Tuloksia voidaan jalostaa monin eri tavoin ja tulokseksi saadaan mm. erilaisia taulukoita ja kuvaajia.

Saimme pystytettyä patenttitietokannan joka toimii erillisessä serverissä LUT:n tiloissa. Tietokannasta saadaan nopeasti paljon tietoa halutuilla hakuehdoilla ja riittävän hyvin varustellun palvelinlaitteiston hankkiminen varmistaa monen yhtäaikaisen käyttäjän tietokantahaut sekä mahdolliset fyysiset laajennukset, kuten prosessorin tai muistin lisäämisen.

LÄHDELUETTELO

[1] Derek J. Balling, Jeremy Zawodny. High Performance MySQL. 2004 O'Reilly. ISBN: 0-596-00306-4

[2] W. Jason Gilmore. Beginning PHP and MySQL 5, Second Edition From Novice to Professional. 2006 Apress. ISBN: 978-1-59059-552-7

[3] Thomson Reuters. Using Patent Citations to Enhance your Subject Searching.

[Verkkodokumentti]. [Viitattu 20.10.2009]. Saatavissa:

http://science.thomsonreuters.com/support/patents/dwpioref/reftools/searchtips/searchtip_oct

[4] World Intellectual Property Organization. International Patent Classification. Eight Edition. 2006. Volume 5 Guide. ISBN 92-805-1442-3

[5] Patents statistics for policy decision making. Venice, Italy. 2007. Konferenssi.

[6] EPO-OECD conference on Patent Statistics for Policy Decision Making. 2006. Vienna. Rob Heijna. James Rollinson.

[7] World Intellectual Property Organization. WIPO Guide to Using Patent Information.

http://www.wipo.int/freepublications/en/patents/434/wipo_pub_1434_03.pdf

[8] CEMI's PATSTAT Knowledge Base. what is PATSTAT. Saatavissa:

<http://wiki.epfl.ch/patstat/whatis>

[9] Bronwyn H. Hall. The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. University of California, Berkeley and NBER.

[10] Microsoft Office Online. Microsoft Access Specifications. Saatavissa:

<http://office.microsoft.com/en-us/access/HP051868081033.aspx>

viitattu 24.11.2009

[11] VT Timo Kivikoskinen. 2008. Teollisoikeudet liiketoiminnassa, tutkimuksessa ja tuotekehityksessä. Luennot 2008.

[12] Bronwyn H. Hall, Adam Jaffe, Manuel Trajtenberg. Market value and patent citations. *The Rand Journal of Economics*, Vol. 36; no. 1, pp. 16-38

[13] Bronwyn H. Hall, Grid Thoma, Salvatore Torrisi. The Market Value of Patents and R&D: Evidence From European Firms. Working Paper 13426. NBER Working Paper Series.

[14] Chen-Lung Chin, Picheng Lee, Hsin-Yi Chi, Asokan Anandarajan. Patent Citation, R&D Spillover, and Tobin's Q: Evidence from Taiwan Semiconductor Industry.

[15] Bronwyn H. Hall. Patent Data as Indicators. University of California at Berkeley, NBER and IFS London. Saatavissa: http://elsa.berkeley.edu/~bhhall/papers/BHH04_WIPO.pdf

[16] Albert G.Z. Hu, Adam B. Jaffe. Patent citations and international knowledge flow: the cases of Korea and Taiwan. NBER Working Paper No. 8528

[17] Teknologia- ja kilpailijatietoa patenttimaisema-analyyseistä. 2008. VTT. Saatavissa: <http://www.vtt.fi/inf/julkaisut/muut/tietopalvelu/patenttialyysiesimerkki.pdf>

[18] World Intellectual Property Organization. International Classifications at WIPO. Saatavissa: http://www.wipo.int/classifications/fulltext/new_ipc/ipcen.html

[19] Espacenet (2009). [WWW-sivut]. Saatavilla: <<http://www.espacenet.com/>>

[20] US-Patent (2009). [WWW-sivut]. Saatavilla: <<http://www.uspto.gov/>>

[21] Freepatensonline (2009). [WWW-sivut]. Saatavilla: <<http://www.freepatensonline.com/>>

[22] World Intellectual Property Organization. National Office Databases. Saatavilla:
http://www.wipo.int/patentscope/en/search/national_databases.html

[23] T. Nikulainen, R Hermans, M. Kulvik. Patent citations indicating present value of the biotechnology business. The Research Institute of the Finnish Economy Discussion papers No. 1048.