

Atte Jäntti

# Power BI -raportti patenttidataalle

Opinnäytetyö

Liiketalouden ammattikorkeakoulututkinto

Data-analytiikka

2024



**Kaakkois-Suomen  
ammattikorkeakoulu**

Tutkintonimike	Tradenomi (amk)
Tekijä/Tekijät	Atte Jäntti
Työn nimi	Power BI -raportti patenttitalle
Toimeksiantaja	Reseptirobotti-hanke
Vuosi	2024
Sivut	34 sivua, liitteitä 0 sivua
Työn ohjaaja(t)	Jarkko Ansamäki, Väinö Toot

## TIIVISTELMÄ

Kaakkois-Suomen ammattikorkeakoulun DataLabin tuottama Python-hakutyökalu hakee lens.org:n yli 150 miljoonan patentin tietokannasta patenteja, käyttäen hakusanoja. Haetut patentit palaavat tietopöytämuodossa, joka ei ole lähestyttävä keskivertokäyttäjälle. Tässä toiminnallisessa opinnäytetyössä on luotu Microsoft Power BI -raportti, joka visualisoi patenttidataa helposti ymmärrettävässä muodossa sekä mahdollistaa patenteista kaikkein kiinnostavimpien seulomisen. Patenttien tutkimisen avustaminen on oleellista tutkimus- ja kehitystyön mahdollistamiseksi.

Opinnäytetyössä määritetään tiedolla johtamisen teorian kautta viitekehys, jolla Power BI -raportin luominen on suoritettu. Tutkimusongelmissa keskitytään datan siivoamiseen, mallintamiseen, visualisoimiseen, sekä lopuksi tiivistämiseen toimeksiantajan Reseptirobotti-hankkeen hyödyksi. Tuotettu Power BI raportti sisältää neljä sivua visualisointeja ja suodattimia patenttien seulontaa varten. Power BI raporttia varten tehdyt askeleet on määritelty tarkasti, jotta jatkokehitys on mahdollista.

Tuotoksen avulla visualisoitua patenttidataa liittyen ekologiseen betonin valmistukseen on tutkittu käyttäen kuvauksellista poikkileikkausaineisto tutkimusta, joka on kvantitatiivinen tutkimusmenetelmä. Tämän tutkimuksen tavoitteena on vastata toimeksiantajaa kiinnostaviin kysymyksiin patenttidataa tiedoista, kuten patenttien määrästä, niiden julkaisumaista, julkaisumäärän kehityksestä aikaan verraten, sekä patenttien jakautumisesta patenttiluokituksiin. Tämän lisäksi, raportin suodattimien avulla patenttidatasta on seulottu toimeksiantajaa eniten kiinnostavat patentit, sekä esitetty kuinka paljon seulominen auttoi rajaamaan patenttidatasta pois epäoleellista tietoa.

**Asiasanat:** patenti, Power BI, tieto johtaminen, business intelligence

Degree title	Bachelor of Business Administration
Author (authors)	Atte Jäntti
Thesis title	Power BI -report for patent data
Commissioned by	Reseptirobotti-project
Time	2024
Pages	34 pages, 0 pages of appendices
Supervisor	Jarkko Ansamäki, Väinö Toot

## ABSTRACT

The Python search tool produced by the DataLab of the Southeastern Finland University of Applied Sciences retrieves patents from the lens.org database of over 150 million patents using keywords. The retrieved patents are returned in tabular form, which is not easily accessible to the average user. In this functional thesis, a Microsoft Power BI report was created to visualize patent data in an easily understandable format and to facilitate the screening of the most interesting patents. Assisting in exploring patents was essential for enabling research and development work.

Through the theory of data management, a framework for creating the Power BI report was established in the thesis. The research focused on cleaning, modeling, visualizing the data, and finally summarizing it to benefit the client's Recipe Robot project. The produced Power BI report contained four pages of visualizations and filters for patent screening. The steps taken for the Power BI report were defined precisely to enable further development.

Descriptive cross-sectional data regarding environmentally friendly concrete production-related patent data was studied using qualitative research methods. The research aimed to answer questions of interest to the client regarding patent data, such as the number of patents, their publication countries, publication trends over time, and the distribution of patents into patent classifications. In addition, using the report's filters, the most exciting patents were screened from the patent data, and it was shown how much this screening helped to filter out irrelevant information.

**Keywords:** patent, Power BI, data management, business intelligence

# SISÄLLYS

1	JOHDANTO.....	5
1.1	Työn taustaa ja tilaajan esittely.....	5
1.2	Keskeisiä käsitteitä .....	6
1.3	Patentit ja niiden luokitukset .....	7
2	BUSINESS INTELLIGENCE ELI BI.....	8
2.1	BI-työkalut ja ohjelmistot.....	8
2.2	Power BI raportin luomisen työkaluna .....	9
3	TIETOJOHTAMINEN.....	9
4	TUTKIMUKSEN TAVOITTEET JA MENETELMÄT .....	11
4.1	Toiminnallinen opinnäytetyö .....	11
4.2	Opinnäytetyön tutkimusmenetelmät.....	12
4.3	Toimeksiantajan patenttirajaus .....	13
5	POWER BI -RAPORTIN VALMISTELUT.....	14
5.1	Datan muoto .....	14
5.2	Datan tuominen ja siivoaminen.....	15
5.3	Datan mallintaminen .....	16
5.4	Raportin rakenne .....	18
6	TUTKIMUKSEN TUOTTAMA POWER BI -RAPORTTI .....	19
6.1	Patents Overview -sivu .....	20
6.2	Patent Classifications Tree -sivu.....	23
6.3	Patents Table -sivu .....	25
6.4	Patent Filters -sivu .....	26
7	TUOTOKSET.....	27
7.1	Tuotokset ja niiden tulkinta Reseptirobotti-hankkeelle .....	28
7.2	Ehdotukset jatkokehitykselle ja oma pohdinta .....	31
	LÄHTEET.....	33

## 1 JOHDANTO

Nykymaailma tuottaa loppumattomia virtoja dataa. Ongelma ei ole onko tietoa, vaan onko tiedon selventämiseen, siivoamiseen ja esittämiseen puitteet ja taidot. Patenteja tehdään jatkuvalla virralla kaikenlaisista asioista, sekä patenttien tarjoamat mahdollisuudet ja rajoitukset vaikuttavat suuresti yritystoimintaan. Yrityksille ja organisaatioille ongelma on kaikkien olennaisten patenttien löytäminen sekä esittäminen helposti ymmärrettävässä muodossa, jotta oma tuotekehitys voi toimia niiden luomissa raameissa eikä vahingossa rikkoa patenttilakeja.

Patentit ovat sekä estävä ja mahdollistava tekijä tutkimus- ja kehitystoiminnassa. Tutkimus- ja kehitystoimintaa kartoittavassa analyysissä selvisi, että 33 % kotimaisista patenttihakemuksista oli epäonnistunut jo olemassa olevan patentin vuoksi, joka koski samaa osa-aluetta, aihetta tai keksintöä (Neuvonen ym. 2010). Patenttikonfliktien välttämisen helpottaminen on yksi syy patenttidataan seulomisen helpottamisen tärkeydelle, sekä tämän tutkimuksen päätaivoite.

Lens Org -organisaatiolla on yli 153 miljoonan patentin tietokanta (Lens Org 2024), jota hyödyntäen Xamkin DataLab on luonut hakutyökalun, joka noutaa annettujen määritelmien perusteella patenttitietoja. Nämä tiedot palaavat tietotaulukko muodossa, jota voi jatkokehittää ja työstää. Kaikki hakutyökalun käyttäjät eivät kuitenkaan osaa seuloa suuria määriä dataa, luoda visualisointeja, taikka suodattaa palanneista tiedoista aiheen ulkopuolisia patenteja pois. Tämän vuoksi tämän tutkimuksen tarkoitus on luoda selkeä raportti, jossa haetut tiedot ovat helposti suodatettavissa, visualisoituina, tarkasteltavissa, helposti ymmärrettävässä ja siistissä muodossa.

### 1.1 Työn taustaa ja tilaajan esittely

Tämän tutkimuksen toimeksiannon juuret alkoivat Dataamo-hankkeen toimeksiannosta data-analytiikan projektityökurssilla, jossa opiskelijat Hannu Illikainen ja Jyrki Kymäläinen loivat Power BI -raportin visualisoimaan e.m. hanketta kiinnostavaa patenttidataa. Tässä kurssiprojektissa patenttidata oli hankittu pyytämällä dataa .csv -tiedostomuodossa Lens-organisaatiolta sähköpostilla.

Projektia lähti jatkokehittämään DataLab. DataLab aloitti toimintansa vuonna 2020, tavoitteena tukea data-analytiikan opiskelijoita, Xamk-organisaatiota, sekä yhdistää niitä yrityselämään (DataLab 2024). Aiemman kurssiprojektin tuotoksia jatkokehitettiin DataLabin laboratoriomestarin Väinö Tootsin toimesta, kun hän loi Python-kielellä hakutyökalun, jolla voi hakea patenttidataa Lens-organisaation rajapinnasta hyödyntäen hakusanoja.

Tämän tutkimuksen toimeksiantaja on Kaakkois-Suomen ammattikorkeakoulu (Xamkin) Reseptirobotti-hanke. Reseptirobotti-hanke on vuonna 2024 alkanut hanke, joka tähtää kehittämään tuotekehitystä käyttäen julkaistua tietoa hiilineutraalin Suomen tieteenalalla (EURA2021 2024). Tässä opinnäytetyössä hanketta kiinnostaa patenttivalikoiman tutkiminen opinnäytetyön luomaa Power BI raporttia käyttäen. Patenttivalikoima on haettu käyttämällä listaa avainsanoja, joista lisää luvussa 1.2.3 Toimeksiantajan patenttirajaus.

## 1.2 Keskeisiä käsitteitä

*Data*: Suuri määrä järjesteltyä tietoa. Usein tietopöytä tai csv muodossa.

*csv -tiedosto*: Tiedostoformaatti, johon voi tallentaa suuria määriä tietoja, joista jokainen erillinen tietue on eroteltu erikoismerkillä.

*Tietue*: Yksi solu tai lokero tietoa data erässä.

*Tietotaulu*: Taulukkorakenne, joka sisältää tietoa tai dataa. Sisältää rivejä tietoa ja sarakkeita tätä tietoa määrittäviä muuttujia.

*Business Intelligence (BI)*: Tiedon rakenteellista käyttöä yritysmaailman päätöksenteossa. Usein myös käytetään viitatessa BI-sovelluksiin.

*Microsoft Power BI*: Microsoftin julkaisema sovellusratkaisu Business Intelligence tehtävien suorittamiseksi.

*Relaatiotietokanta:* Relatiivisissa tietotaulujen välille luodaan yhteyksiä, jotka määrittävät miten ne voivat mm. suodattaa toisiaan tai miten niiden tietoja voidaan yhdistää.

*Kardinaliteetti (tietokannoissa):* Mallinnuksessa käytettävä termi, joka määrittää tietotaulujen väliset suhteet. Esim. Mihin suuntaan tietotaulut voivat suodattaa, vain yhteen vai voiko molemmat suodattaa molempia.

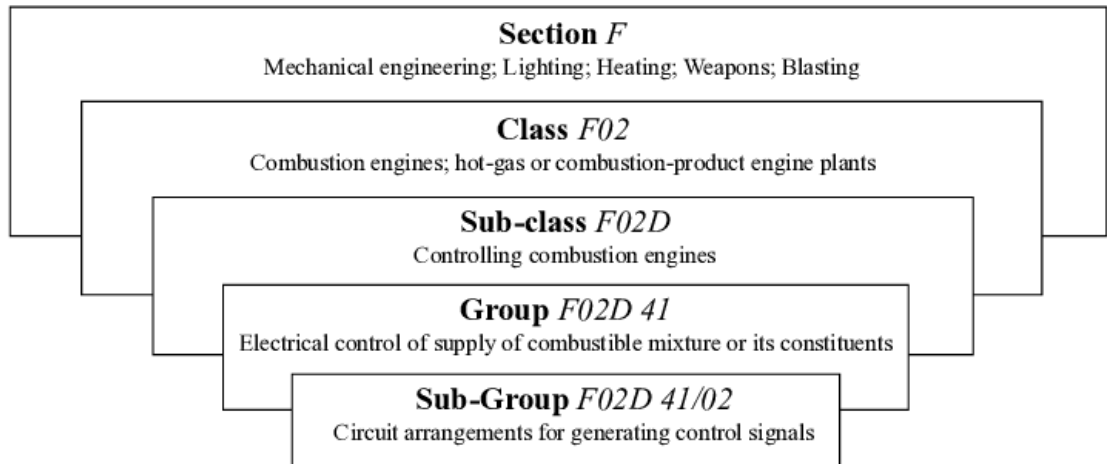
*Relatiomalli:* Teoria, jonka pohjalta tietokannan mallinnus on suoritettu. Esim. Tähtiskeema. Määrittää toimintaperiaatteet uusien tietotaulujen luonnissa, sekä miten niitä voi yhdistää toisiin tietotauluihin.

*Avainarvo:* Tietotaulun sarake, jonka arvo määrittää miten se voi luoda yhteyksiä toisiin tietotauluihin. Faktapöydissä jokainen avainarvo tietue on uniikki, kun taas ulottuvuuspöydissä monta arvoa voi edustaa samaa avainarvoa.

*API-rajapinta:* Englannin sanoista Application Programming Interface. Tarkoittaa verkkosivun käyttöliittymää, jota voidaan kutsua ohjelmoinnin kautta, ilman että joudutaan avaamaan graafinen käyttöliittymä. Käytetään esim. tietojen noutamiseen.

### **1.3 Patentit ja niiden luokitukset**

Patentit arkistoidaan sisältäen monenlaista meta- ja julkaisudataa. Näihin kuuluu mm. patenttien julkaisupäivämäärät, tietokantaan lisäyksen päivämäärät, julkaisumaa, julkaisijan tietoja, patentin luokitus, sekä muuta luokittelevaa dataa. Patentit luokitellaan hierarkkisesti (kuva 1). Laakein luokitus on Section, jonka alla on alaluokitukset Class, Sub-class, Group & Sub-Group.



Kuva 1. Patenttiluokitukset ja niiden hierarkia (Risch ym. 2020)

Tässä opinnäytetyössä keskitytään patenttien määrään, julkaisuaikoihin, patenttiluokituksiin, sekä kuvauksiin ja abstrakteihin, joista saa silmäiltyä tarkempaa tietoa valitusta patentista. Patenttiluokituksiin keskitytään erityisen paljon, tavoitteena mahdollistaa tuotetussa raportissa patenttien selaaminen ja rajaaminen niiden luokitusten avulla. Näin raportin loppukäyttäjät pystyvät saamaan paremman ymmärryksen haettujen patenttien sisällöistä, sekä rajaamaan pois sellaiset patenttiluokitukset, jotka eivät heitä kiinnosta.

## 2 BUSINESS INTELLIGENCE ELI BI

Business Intelligence (tästä eteenpäin BI) kuvaa yritysmaailman päätöksentekoa tukevaa datan hyödyntämistä (TableAU 2024). Sen voi määritellä tietojohdamisen kanssa samaan alaan, tai kenties tietojohdamisen alakategoriaksi, kun tietojohdaminen kattaa enemmän käsitteitä. BI-termillä viitataan kuitenkin useinten BI-ohjelmistoihin, jotka tarjoavat työkalut BI-tehtävien suorittamiseen.

### 2.1 BI-työkalut ja ohjelmistot

BI-ohjelmistosta riippumatta, niillä tähdätään pääsijaisesti ratkaisemaan samoja ongelmia. Niitä yhdistää kyky tuoda sisään datalähteitä, luoda niiden välille yhteyksiä, korjata mahdollisia datavirheitä, sekä luoda erilaisia analyyseja ja visualisointeja tuodun datan perusteelta. Ohjelmistot eroavat mm. käyttöliittymän, lisenssien hintojen, datalähteiden hallinnan periaatteiden, sekä visualisointien luomisen ja visualisointi valikoiman laajuuden osa-alueilla.



Power BI on Microsoftin vuonna 2014 julkaisema ohjelmisto, joka on tarkoitettu pääasiallisesti yritysten datan hallintaan, tulkintaan ja visualisointiin. Sillä voi luoda mittaristoja ja raportteja, jotka päivittyvät reaaliajassa niiden datalähteiden päivittyessä uusilla tiedoilla. (Microsoft 2024.)

Power BI:lla on varteenotettavia kilpailevia BI-ohjelmistoja, kuten Tableau tai Snowflake. Niilläkin olisi voinut toteuttaa tämän tutkimuksen tuotokset, mutta koska Xamk organisaationa on sitoutunut Microsoft-ohjelmistoihin, Microsoftin oma BI-ohjelmisto oli paras valinta etenkin jatkokehitysmahdollisuuksien kannalta. Näin tämän opinnäytetyön tuotoksia voidaan linkittää muihin koulun tietokantoihin, tai raportin visualisointeja voidaan lisätä muihin raportteihin.

## **2.2 Power BI raportin luomisen työkaluna**

Power BI palvelu koostuu kahdesta osasta: tietokonesovelluksesta ja pilvipalvelusta. Tietokonesovellus Power BI desktop on datalähteiden yhteyksiä ja raporttien luomista varten. Sen lisäksi on Power BI -pilvipalvelu, johon pääsee internetselaimella. Pilvipalvelussa voi nähdä ja rajoitetusti muokata raportteja, sekä tulostauluja (Dashboard), kun taas Desktop-versio on datatyöskentelijän pääalusta, josta sisältöä julkaistaan pilvipalveluun. (Microsoft 2024).

Tässä toiminnallisessa opinnäytetyössä tähdätään tuottamaan laadullinen ja visuaalisesti vetoava Power BI -raportti. Power BI desktop -versiolla tullaan yhdistämään sekä siivoamaan patenttidatalähteet, jonka jälkeen luodaan raportti sisältäen suodattimia, visualisointeja ja selitteitä ohjaamaan loppukäyttäjää ymmärtämään raportin käyttöä. Kun raportti on valmis, se julkaistaan Power BI:n pilvipalveluun, josta loppukäyttäjät voivat käyttää sitä.

## **3 TIETOJOHTAMINEN**

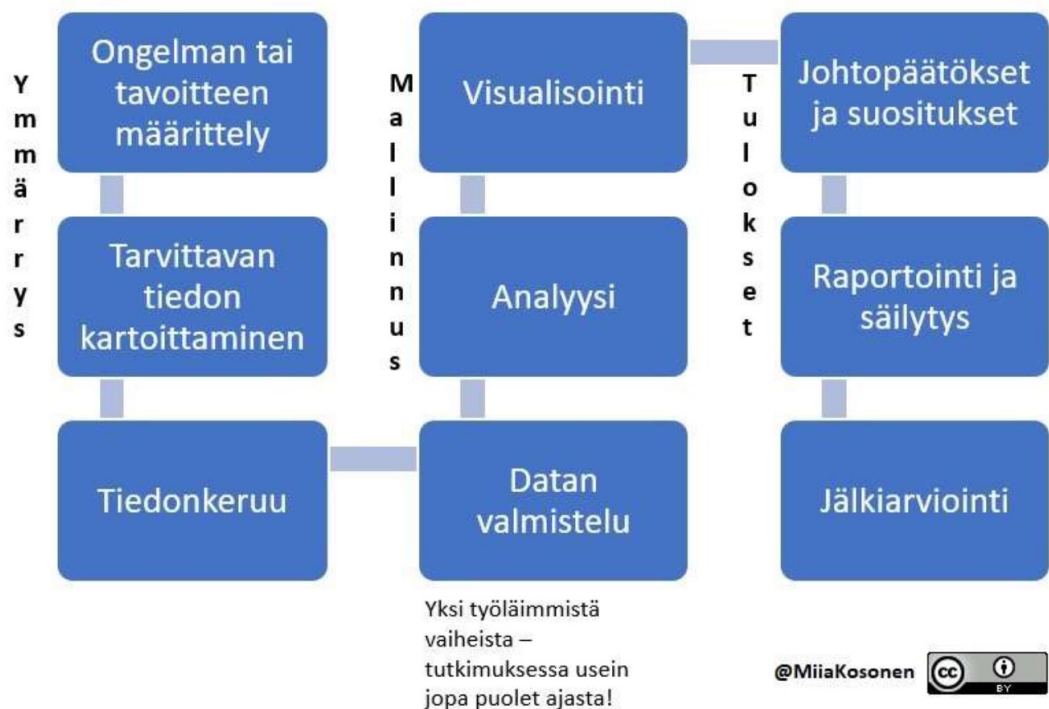
Julkaisussaan Tiedolla johtamisen käsikirja, Kosonen (2019) kiteyttää tiedolla johtamisen olevan rakenteellinen lähestymistapa tiedon hallitsemiseen ja käyttämiseen, sekä osa suurempaa tietojohdamisen tieteenalaa. Tiedolla johtaminen koostuu mm. tiedon tuottamisesta, hallitsemisesta, säilömisestä ja analysoimisesta, sekä myös sen tiedon käyttämisen osaamisesta uuden arvon tuottamiseksi (Kosonen 2019).

Patenttitiedon hyödyntäminen yritystoiminnassa vaatii tiedolla johtamista. Jotta valtavista määristä patenteja voi luoda koonteja ja tulkittavia lopputuotoksia, se vaatii osaamista datan keräämisen, korjaamisen, hallitsemisen, esittämisen ja analysoimisen osa-alueilta.

Tiedolla johtaminen sisältää kaikenlaisia työtehtäviä, mutta ne voidaan jakaa pääosin kolmeen kategoriaan: ymmärrykseen, mallinnukseen ja tuloksiin.

Tämä opinnäytetyö seuraa rakenteeltaan tiedolla johtamisen prosessin askeleita (kuva 2).

## Tiedolla johtamisen prosessi



Kuva 2. Tiedolla johtamisen prosessin askeleet (Kosonen 2019)

Tässä tutkimuksessa, kuvassa 2 eritellyistä askelista ymmärryksen puolelta vaiheet ”ongelman tai tavoitteen määrittely” ja ”tarvittavan tiedon kartoittaminen” ovat huomioitu ja tavoitettu luvussa 4 Tutkimuksen tavoitteet ja menetelmät. Kyseisessä luvussa eritellään tutkimuksen tavoitteet, tutkimusongelmat ja tutkimuskysymykset, sekä mikä tieto toimeksiantajaa kiinnostaa. Tiedonkeruu-

askel ei ole tarpeellinen, sillä tätä tutkimusta varten oli kerätty tieto Python-hakutyökalulla.

Työläin tiedolla johtamisen prosessista ja valtaosa tämän opinnäytetyön työtaakasta keskittyy kuvassa 2 ilmeneviin mallinnuksen vaiheisiin. Mallinnuksen vaiheet ”datan valmistelu” ja ”analyysi” eritellään luvussa 5 Power BI raportin valmistelut, jossa datalähteet tuodaan sisään Power BI:hin, ne korjataan virheistä ja ongelmista, datan yhteydet luodaan mallinnuksen avulla, sekä suunnitellaan tuotettavan raportin rakenne. Luvussa 6 Tutkimuksen tuottama Power BI luodun raportin sisältö esitellään.

Tähän loppuu opinnäytetyön osuus tuotoksen luomisesta, jolloin siirrytään kuva 2 osuuteen ”tulokset”. Opinnäytetyö itsessään sisältää ”tulokset” vaiheen kaikki askeleet, sillä opinnäytetyössä raportoidaan otetut askeleet, tuotokset, johtopäätökset ja jatkokehityksen ehdotukset, se esitetään muille, työ arvioidaan ja lopuksi arkistoidaan ja julkaistaan.

## **4 TUTKIMUKSEN TAVOITTEET JA MENETELMÄT**

### **4.1 Toiminnallinen opinnäytetyö**

Tämä opinnäytetyö on toiminnallinen ja sen päätavoite on luoda tuotos. Suurin osa toiminnallisista opinnäytetöistä luo tuotoksen toimeksiantajan käytettäväksi (Vilkkä & Airaksinen 2003, 9.). Tuotos, joka tutkimuksen ohella on luotu, on Power BI -raportti esittämään Python-työkalun noutamaa patenttidataa, antaen helposti ymmärrettävän yleiskatsauksen haetusta datasta.

Power BI -raportin toteutuksessa tulee huomioida esityksen helposti ymmärrettävyys sekä visuaalinen vetoavuus. Lisäksi jatkokehitys tulee huomioida niin, että raportin skaalautuvuus ja toimivuus muunkin patenttierän kanssa on otettu huomioon.

Opinnäytetyön keskiössä ovat tuotoksen kehittämistehtävät:

1. Siivoa data ja luo datayhteydet eli mallinnus Power BI -raporttia varten.
2. Suunnittele ja luo raportin sisältö, sekä julkaise se Power BI -pilvipalveluun.
3. Testaa raporttia toimeksiantajan määrittämää patenttidataa hyödyntäen.

Luku 5 ”Power BI -raportin valmistelu”, jossa eritellään otetut askeleet datan tuomiseen Power BI:hin, sen siivoamiseen, mallinnukseen, sekä suunnitelma raportin sisällön luomiseksi, sisältää ensimmäisen kehittämistehtävän. Luku 6 ”Tutkimuksen tuottama Power BI -raportti”, jossa eritellään raportin sisältö ja toiminnallisuudet, sisältää toisen kehittämistehtävän. Luku 7 ”Tuotokset ja raportin tulkinta”, jossa raporttia testataan käyttämällä toimeksiantajan määrittämää patenttidataa, sisältää viimeisen kehittämistehtävän.

#### **4.2 Opinnäytetyön tutkimusmenetelmät**

Osa tätä opinnäytetyötä on kvantitatiivinen tutkimus koskien toimeksiantajaa kiinnostavaa patenttidataa. Kvantitatiivisessa tutkimuksessa keskitytään määrälliseen ja tilastoitavaan tietoon (Tietoarkisto 2024). Valittu tutkimusmenetelmä on kuvauksellinen poikkileikkausaineisto tutkimus, jossa annetun patenttieron muotoa selvitetään ja erotellaan.

Patenttidatan muodon erittelyn jälkeen, pitäen toimeksiantajan kiinnostuksen kohteet mielessä, tähdätään suodattamaan epäolennaisia patenteja pois patenttierästä. Tämän jälkeen eritellään, kuinka paljon suodatus vähensi patenttien määrää, minkä tarkoituksena on esittää toimeksiantajalle opinnäytetyön tekemän tuotoksen hyödyllisyys patenttien seulomisessa.

Suorittaessa kuvauksellista poikkileikkaustutkimusta luvussa 7 ”Tuotokset”, vastataan seuraaviin tutkimuskysymyksiin:

1. Montako patenttia hakusanoilla löytyi?
2. Mistä maista tai organisaatioista on tullut eniten patenttijulkaisuja?
3. Miten patentit jakautuvat patenttiluokituksiin: mitä luokituksia on eniten?
4. Miten patenttien julkaisujen tahti on kehittynyt vuosien aikana?
5. Patenttien suodattamisen jälkeen, paljonko aiemmin eriteltyjen patenttijakaumien määrät ovat laskeneet?

### 4.3 Toimeksiantajan patenttirajaus

Toimeksiantaja Reseptirobotti-hanke on antanut listan hakusanoista, jota on käytetty rajaamaan Python-työkalulla tehdyn patenttidata haun. Hakusanat ja-  
kautuvat kolmeen kategoriaan: yleiset-, koostumus- ja ominaisuushakusanoi-  
hin (taulukko 1).

Taulukko 1. Reseptirobotti-hankkeen antama patenttirajaus

Hakusanan kategoria	Hakusanat
Yleisiä hakusanoja	Low carbon concrete Green concrete Eco concrete Sustainable concrete Carbon footprint reduced / reduction Low carbon footprint
Koostumushakusanoja	Cement reduction Replacement of Portland cement Low carbon cement alternatives Alkali activation Geopolymer Mineral compounds Calcined clays Fly ash Blast-furnace Slag GGBS & GGBS cement Ground Granulated Blast-furnace Slag Recycled aggregates / aggregate recycling
Ominaisuushakusanoja	Compression strength Development of initial strength Strength development Durability Water absorption Salt freeze resistance Freeze thaw resistance

Efflorescence  
Sulphate resistance

Tuotokset ja raportin tulkinta -luvussa tätä patenttirajausta käytetään pinnallisen analyysin suorittamiseen. Osa tämän opinnäytetyön tuottavaa osuutta on tehdä kiteytys raportin esittämistä löydöksistä, pitäen mielessä mikä toimeksiantajaa voisi kiinnostaa. Tutkien patenttirajauksen hakusanoja, toimeksiantajaa kiinnostavat patentit voidaan tiivistää liittyvän ekologiseen betoniin ja sen ominaisuuksiin liittyviin patentteihin.

## 5 POWER BI -RAPORTIN VALMISTELUT

Ennen kuin raportin voi luoda Power BI desktopissa, on useita askelia, jotka täytyy ottaa ensin. Niihin kuuluu datan tuominen Power BI:hin, datan siivoamisen askeleiden määrittäminen käyttäen Power BI:n Power Query -toimintoa, sekä tuodun datan yhdistäminen toisiinsa, eli mallintaminen.

Kun edellä mainitut askeleet ovat suoritettu ja ne toimivat virheettää, voidaan luoda tarkkoja ja monipuolisia raportteja. Jos dataa ei ole siivottu oikein, visualisoinnit voivat näyttää todella vääristäviä tuloksia, joten datan oikeamuotoisuuden varmentaminen on erittäin tärkeää. (Microsoft 2023b)

### 5.1 Datan muoto

DataLabin hakutyökalu noutaa patenttitietoja, syöttäen ne palvelimella olevaan tietokantaan. Nämä tiedot noudetaan Power BI:hin, formatoidaan oikeaan muotoon sekä siivotaan mahdollisista ongelmista käyttäen Power Query -työkalua, joka on osa Power BI -ohjelmistoa.

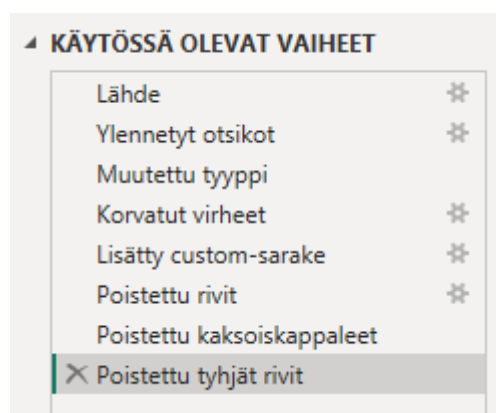
Kun tuotu data on siistitty, se mallinnetaan, jotta eri tietopöydät osaavat kommunikoida toistensa kanssa. Tämä tehdään pitäen mielessä raportin mahdollinen jatkokehitys opinnäytetyön jälkeen.

Mallinnuksen jälkeen luodaan Power BI -raportti. Raportti on lopputuote, jota käyttäjä pääsee tulkitsemaan, suodattamaan ja selaamaan. Raportti sisältää useita visualisointeja antamaan ymmärryksen haetusta patenttidatasta, tekstikenttiä selittämään raportin eri osa-alueita sekä suodatusvaihtoehtoja, joilla käyttäjä pääsee rajaamaan patenttidataa häntä eniten kiinnostaviin patentteihin.

## 5.2 Datan tuominen ja siivoaminen

Kun Power BI:hin tuodaan dataa lähteistä, niitä täytyy lähes aina siivota tai uudelleen formatoida, jotta eri lähteet toimivat yhdessä, sekä jotta Power BI ymmärtää minkä muotoista tuotu data on. Data siivotaan pääosin Power Query -editorilla, mutta datan siivoamisen askelia voi myös suorittaa, kun dataa kerätään itse lähteeseen. Jos on mahdollista, onkin älykästä tähdätä keräämään tieto jo valmiiksi helposti käytettävässä muodossa, jolloin välttyy lisäaskelilta sekä mahdollisilta kompromisseilta ja ongelmilta myöhemmissä datan siivoamisen vaiheissa.

Tämän projektin datalähde vaati useita siivousaskelia. Koska datalähde koostui API-rajapinnan kautta kerätyistä datasta, siinä oli ongelmia mm. puuttuvien tietojen, väärässä muodossa olevan tiedon, sekä väriin soluihin asetettujen tietojen kanssa. Jokaista otettua datan siivoamisen askelta ei eritellä, vaan pääpiirteet.



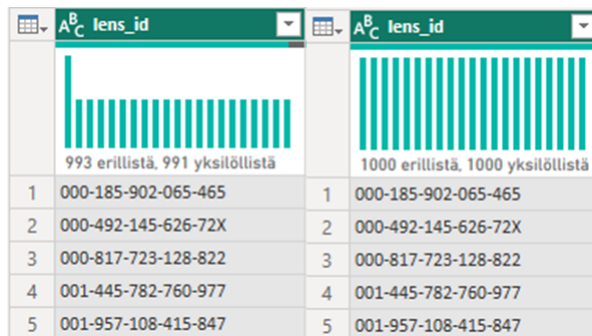
Kuva 3. Power Queryn datan siivoamisen askeleet / operaatiot.

Kuvassa 3 näkyy Power BI:n Power Query -editorin näkymän sivupalkki, joka esittää sillä hetkellä valitun datalähteen otetut siivousaskeleet. Power BI ei muokkaa itse datalähdettä, joten tehdyt siivoukset koskevat vain itse Power BI -projektia. Näitä askelia voi poistaa, siirrellä tapahtumaan eri järjestyksessä tai muokata sitä mitä ne tekevät.

Siivouksen askeleet ovat tärkeä suunnitella niin, että projekti pystyy skaalautumaan. Jos valitaan virheelliset rivit esim. rivimäärän perusteella, tämä siivous-

askel ei toimikaan, jos datalähteeseen lisätään tai sieltä poistetaan rivien määrää. Projektin skaalautuvuuden vuoksi on erittäin tärkeää luoda siivoamisen eri vaiheet niin, että käytetään laakeampia datan varmentamisen askelia, kuten tarkastetaan, onko tietyn sarakkeen rivien tiedoissa oikea määrä kirjaimia, tai onko niissä ainoastaan numeroita. (process.st 2024.)

Jokaiselle eri datalähteelle tehtiin nämä samat siivoamisen operaatiot: ylentämään ensimmäisen rivin otsikoksi, määrittämään sarakkeiden datamuodot, korvaamaan virheellistä tietoa tyhjällä tiedolla, sekä poistamaan tyhjiä tai väärää tietoa sisältäviä avainkenttärivejä. Etenkin avainkenttäsarakeet eivät saa missään tapauksessa sisältää tyhjiä tietoja tai virheellistä tietoa. Avainkenttäsarakeet ovat vaatimus mallinnuksen onnistumista varten. Jos ne sisältävät edes yhden virheen, mallinnus ei toimi ollenkaan. (Microsoft 2024).



Kuva 4. Power Queryn koostenäkymä sarakkeen muuttujien muodosta.

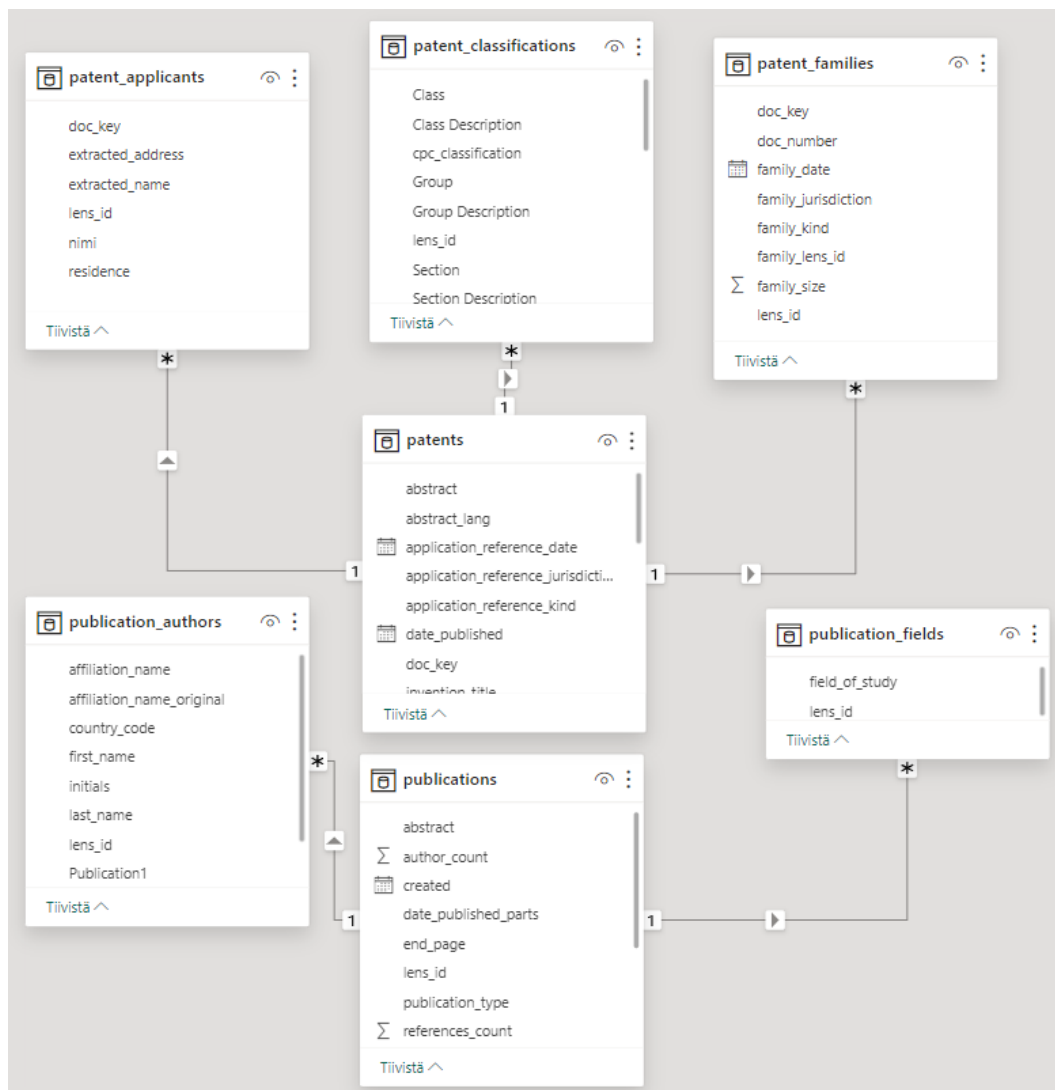
Kuvassa 4 näkyy yhden datalähteen avainkenttäsarakeen koonti *ennen* siivouksen askelia (vasen), sekä siivouksen *jälkeen* (oikea). Ennen siivoamisen askelia, avainkenttäsarake sisältää usean muotoista tietoa. Projektin alkuvaiheissa tämä esti mallinnuksen toimivuuden kokonaan, ennen kuin virheellinen tieto havaittiin ja näin voitiin korjata tai poistaa.

### 5.3 Datan mallintaminen

Kun datan on siivottu ja määritetty oikeisiin muotoihin, sitä voidaan käyttää mallintamiseen. Se tarkoittaa eri datalähteiden yhdistämistä toisiinsa avainsarakkeiden avulla. Yhdistäminen on välttämättömyys, jos eri datalähteitä halutaan käyttää yhdessä, esim. samoissa visualisoinneissa.



On suositeltavia, hyväksi todettuja käytänteitä miten tämän toteuttamiseen. Power BI:ssa parhaiten toimii tähti- tai toisella nimellä lumihutale -malli, jossa kaikki tietopöydät määritellään joko fakta- tai dimensiotietopöydiksi. Faktapöytä tai faktapöydät sisältävät usein valtavia määriä rivejä, kun taas dimensiotietopöydät sisältävät vähemmän tietoa, jota käytetään määrittelemään ja suodattamaan faktapöytien suurta tietomäärää. Näin projekti säilyy helposti skaalautuvana, tietokoneen resurssien hallinnan kannalta tehokkaana, sekä selkeänä ymmärtää riippumatta mallinnuksen tulkitsijasta. (Microsoft 2023c).



Kuva 5. Projektissa käytetty Power BI -mallinnus.

Kuvassa 5 näkyy projektin Power BI -mallinnuksen toteutus. Se perustuu kahden faktapöytään: patents & publications, joihin on yhdistetty dimensiotietopöytiä. Tulee kuitenkin huomioida, ettei patenteja ja julkaisuja ole yhdistetty toisiinsa. Tämä on sen vuoksi, koska tietopöytien sisältö on liian eri muotoista, jotta niitä voisi yhdistää toisiinsa järkevällä tavalla.

Kuvasta 5 voidaan havainnoida tietopöytien väliset yhteydet; viivat, joiden keskellä on nuoli, sekä alussa ja lopussa olevan merkin. Viivat edustavat avainkenttien välille tehtyjä yhteyksiä, niissä olevat nuolet suodattautuvuuden suuntaa, sekä viivan alun merkki millainen yhteys on kyseessä.

Power BI:ssa voidaan luoda joko yksi moneen tai monta moneen yhteyksiä. Tämä viittaa avainkenttien lukumäärään: joskus avainkenttä yhdessä tietopöydissä voi viitata useaan avainkenttään toisessa, kun taas toisissa jokainen avainkenttä on täysin yksilöllinen verrattaessa toiseen tietopöytäan. Tämän projektin tähtiskeemassa faktapöydät publications ja patents omaavat yksi moneen yhteydet dimensiotietopöytiin ja suodatus toimii vain dimensiopöytiä kohti.

Kun yhteydet on luotu oikein, se mahdollistaa tietopöytien toimivuuden toistensa kanssa. Tätä voi käyttää useiden tietopöytien suodattamiseen interaktiivisten visualisointien muodossa raporttia käyttäessä. Lisäksi tietolähteitä voi muokata Power Query -toimintoa käyttäen suhteessa toisiinsa, sen sijaan että jokainen tietopöytä olisi ainoastaan yksilö.

#### **5.4 Raportin rakenne**

Raportin rakenne on tärkeä suunnitella alituisesti pohtien ”mitä käyttäjä haluaisi nähdä?”, sekä ”mikä hämmentää käyttäjä?”. Patenttidataan tutustuneena on helppoa unohtaa, ettei loppukäyttäjä ymmärrä kaikkea mitä itse ymmärtää, sekä että tieto pitää esittää hyvin selkeästi. Eli täytyy harkita kohdeyleisö, sekä heidän tarpeensa ja kykynsä (Microsoft 2023d).

Tämän opinnäytetyön raportti sisältää neljä sivua: Patents Overview -sivun, Patent Classification Tree-sivun, Patent Table -sivun sekä Patent Filters -sivun. Raportti on englannin kielellä, sillä raportin loppukäyttäjiin voi hyvinkin kuulua ulkomaalaisia tahoja. Ensimmäinen mitä loppukäyttäjä näkee, on Overview-sivu. Se sisältää perustiedot haetusta patenttidatasta, kuten paljonko patenteja on, mistä maista niitä on julkaistu eniten, milloin niitä on julkaistu, sekä miten julkaisujen määrä on kehittynyt ajan myötä. Patent Classifications Tree -

sivu on kenties tärkein, sillä siellä oleva puudiagrammi auttaa ymmärtämään miten haetut patentit jakautuvat patenttiluokituksiin. Patent Table -sivu sisältää taulukon patenttien linkeistä, nimistä ja abstrakteista, tarkoituksena rajata taulukon sisällön määrää käyttämällä suodattimia. Patent Filters -sivu sisältää suodattimia, jotka auttavat käyttäjää valitsemaan vain heitä kiinnostavat patenttitiedot suodattimilla. Vaikka DataLABin hakutyökalu rajaa hakunsa avainsanoilla, se palauttaa patenteja, jotka eivät liity aihealueeseen, joten loppukäyttäjän oma patenttien rajausta on tärkeää.

Raportin käyttäminen on suunniteltu toimimaan näin: Käyttäjä avaa raportin, näkee Patents Overview -sivun yleiskatsauksen haetuista tiedoista ja alkaa muodostamaan ymmärrystä haetusta patenttidatasta. Hän siirtyy Patent Classifications Tree -sivulle, josta hän tutkii millaisiin patenttiperheisiin haetut patentit jakautuvat. Hän muodostaa mielipiteen siitä, mitkä perheet eivät sisällä hyödyllistä tietoa, sekä mistä löytyy hänelle tärkeimmät tulokset. Hän siirtyy Patent Filters -sivulle, jossa hän rajaa haettuja patenteja sisällyttämään vain hänelle oleellisimmat patentit. Hän palaa aiempien sivujen visualisointeihin, joita hän tutkii lisää ja tekee johtopäätöksiä. Lopuksi käyttäjä tutkii yksittäisiä patenteja niiden nimien ja abstraktien perusteella Patents Table -sivulla.

On myös tärkeää huomioida käyttäjän mahdollisuus nähdä toimiansa aiheuttamat vaikutukset. Jos sivulla on ainoastaan suodattimia eikä visualisointeja, käyttäjä ei voi tietää toimivatko suodattimet, tai tuottavatko ne haluamiaan tuloksia. Sen vuoksi Patent Filters -sivulle on sisällytetty myös visualisaatio.

## **6 TUTKIMUKSEN TUOTTAMA POWER BI -RAPORTTI**

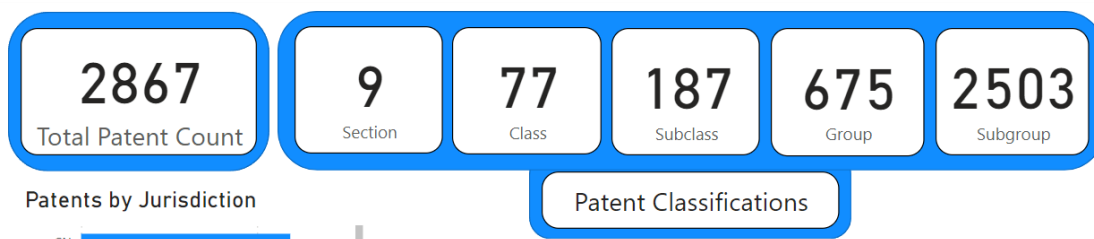
Tämän opinnäytetyön tuottama Power BI -raportti sisältää neljä sivua, joiden sisällöt eritellään tämän luvun alaluvuissa Patents Overview -, Patent Classifications Tree -, Patent Filters -, ja Patent Table -sivu. Raportti on jaettu eri sivuille visualisointien ja suodattimien vaatiman tilan sekä selkeyden vuoksi.

Power BI -raportteja voi suodattaa kahdella tapaa: joko painamalla visualisoinneista eri kohtia, jolloin raportti rajautuu koskemaan valittua tietoa tai erillisillä suodattimilla. Jos toivotaan eri sivujen välistä suodatusta, vaaditaan erillinen

suodatus: visualisoinnin painallukset eivät suodata kaikkia sivuja. Tämän tutkimuksen Power BI -raportissa patenttisuodattimet-sivulla sijaitsevat suodattimet, jotka vaikuttavat kaikkiin sivuihin.

## 6.1 Patents Overview -sivu

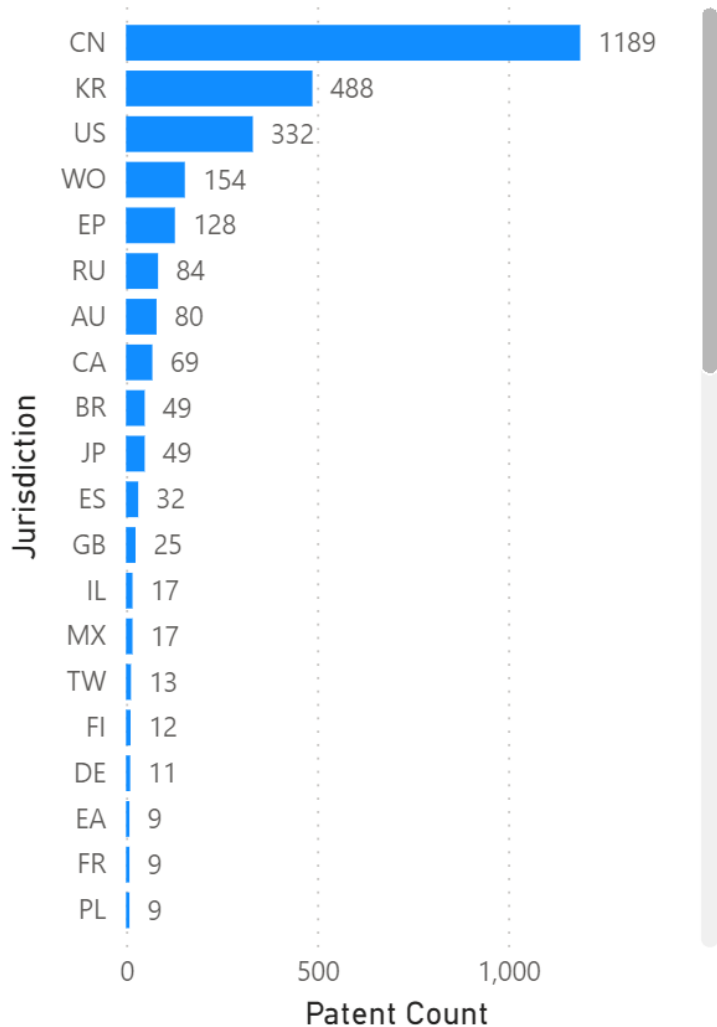
Tämän sivun tarkoitus on antaa nopea yleiskuva haetusta patenttidatasta. Se sisältää patenttien kokonaismäärän, niiden jakautumisen patenttiluokituksiin, eniten julkaisuja tehneet maat, sekä miten patenttien määrä on kehittynyt ajan myötä. Raportin myöhemmällä sivulla olevia suodattimia käyttäessä, tämän sivun katsaus patenttidatasta rajautuu ainoastaan käyttäjää kiinnostaviin patentteihin, jolloin käyttäjän on hyvä palata siihen näkemään muutokset.



Kuva 6. Overview-sivun patenttiluku & patenttiperheisiin jakautuminen.

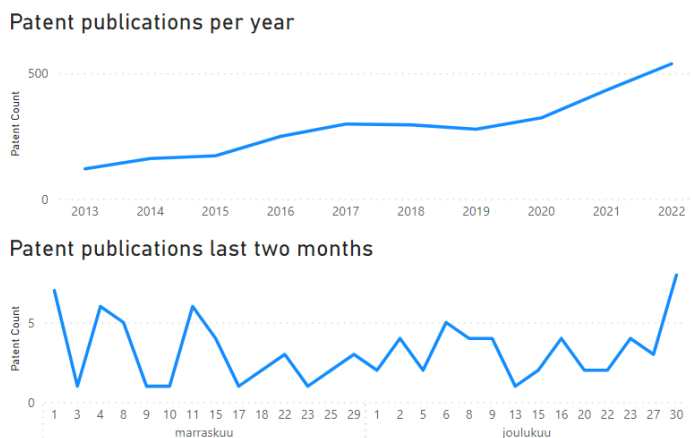
Kuvassa 6 ensimmäiset havainnollistamista tukevat tiedot koskevat patenttien kokonaismäärää sekä patenttiluokituksia. Vasemmalla näkyy patenttien kokonaismäärä sekä oikealla miten patentit jakautuvat patenttiluokituksiin. Patenttiluokituksien tietoja voi tutkia tarkemmin sivulla Patent Classifications Tree.

## Patents by Jurisdiction



Kuva 7. Haettujen patenttien eniten julkaisuja tehneet maat / organisaatiot.

Kuvassa 7 näkyy visualisointi patenttijulkaisuista, järjestelty eniten julkaisuja tehneiden maiden / organisaatioiden perusteella. Julkaisumaa määrittää mm. kuinka helppoa kunkin patentin tietoja on selata ja lukea, sillä esim. Kiinassa julkaistut patentit ovat usein mandariinikiinan kielellä, jota useimmat raportin loppukäyttäjät eivät osaa lukea. Tätä ongelmaa ollaan Python-hakutyökalulla helpotettu tekoälykääntämällä patenttien tietoja englannin kielelle. Julkaisumaa antaa myös kontekstia ymmärtää haettujen patenttien tietoa, sillä joissain alueilla seurataan eri käytäntöjä.

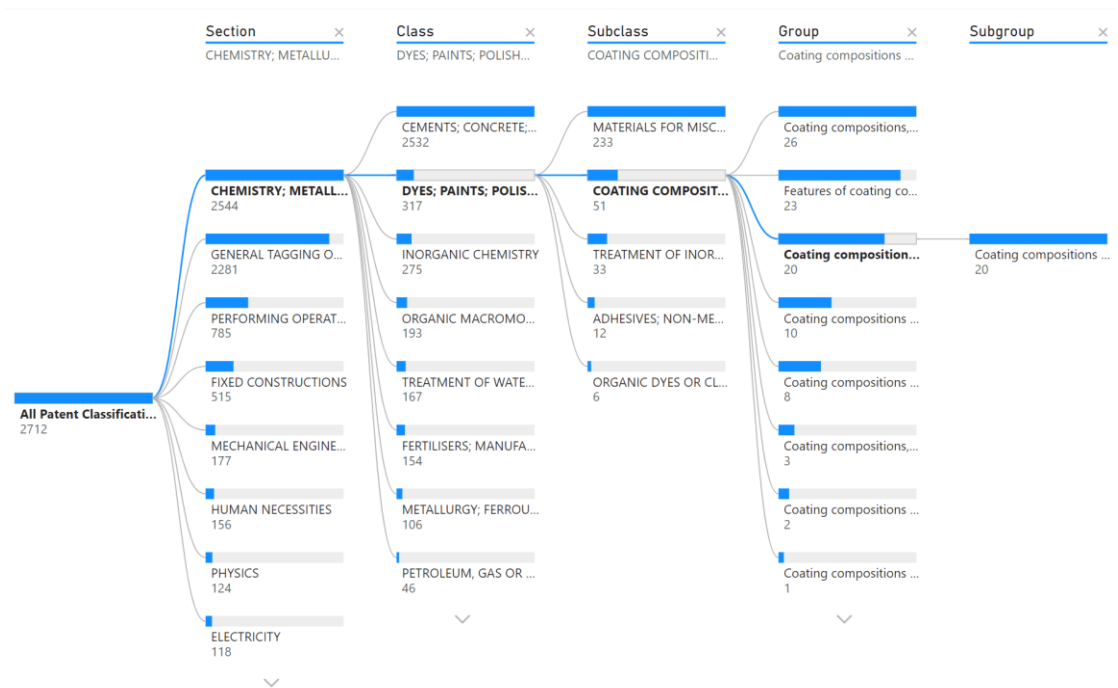


Kuva 8. Haettujen patenttien julkaisumäärien pitkäaikainen- ja lyhytaikainen kehitys.

Kuvassa 8 näkyy patenttijulkaisujen kehitys verrattuna aikaan. Ylemmässä visualisoinnissa verrattuna vuosittaiseen patenttijulkaisujen määrään, toisessa keskittyen kahteen viime kuukauteen. Vuosittaisesta julkaisujen määrästä näkee heti, että patenteja julkaistaan jatkuvasti kasvavalla tahdilla.

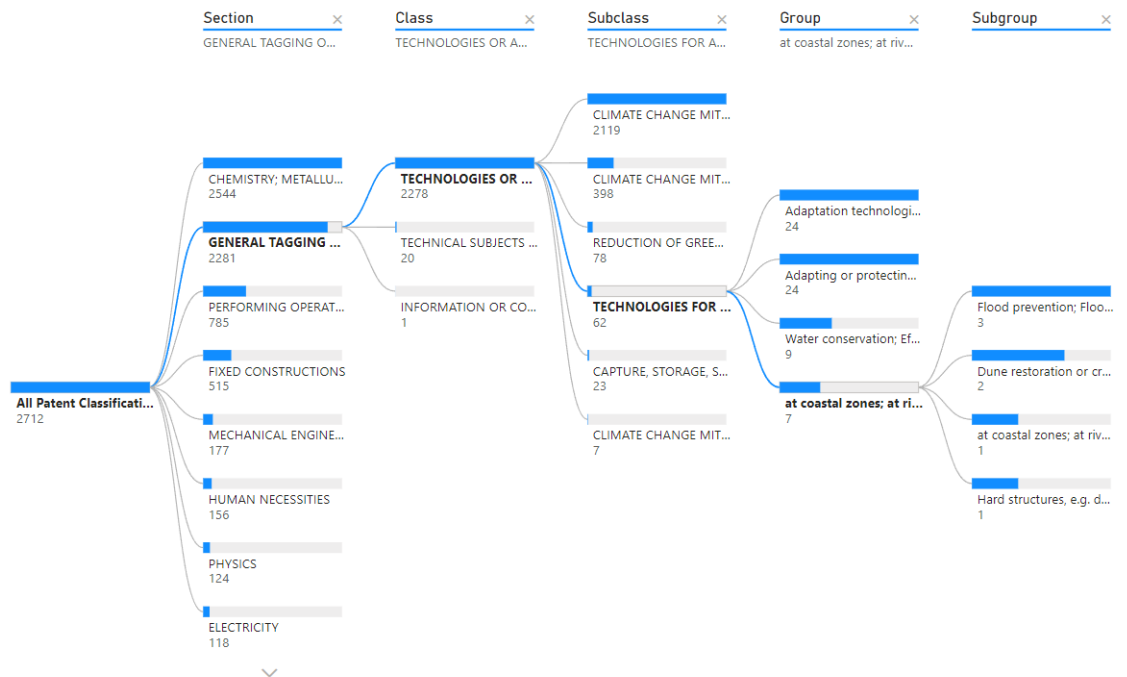
Toimeksiantajaa kiinnostaa patenttien määrän kehittyminen sekä patenttiluokituksiin jakautuminen. Tämän vuoksi kuvan visualisoinnit vuosittaisesta patenttijulkaisujen määrän kasvusta sekä viimeisen kahden kuukauden julkaistuista patenteista on tärkeää sisällyttää raporttiin. Jatkokehityksen näkökulmasta myös muita tahoja varmasti kiinnostaa patenttijulkaisujen kehitys.

## 6.2 Patent Classifications Tree -sivu



Kuva 9. Patenttiluokitusten puudiagrammi -visualisaatio.

Kuvassa 9 näkyy patenttiluokitusten hierarkiaa visualisoiva puudiagrammi. Patentit jaotellaan patenttiluokituksiin, jotka auttavat hahmottamaan ja suodattamaan itselleen oleellisia patenteja. Suurimmalla tasolla on Section, joka lopulta päättyy tarkimpaan määritelmään Subgroup. Puudiagrammin tulokset ovat lajiteltu niiden sisältämien patenttien lukumäärien perusteella, suurimmasta lukumäärästä pienimpään. Selaamalla alaspäin voi kuitenkin nähdä kaikki patenttiluokitukset.



Kuva 10. Esimerkki siitä miten loppukäyttäjä voi muokata ja tutkia puudiagrammia.

Käyttäjä pystyy suodattamaan puudiagrammia klikkaamalla sen tietoja. Tällöin puudiagrammi muuttuu avaamalla eri haaran. Kuvia 9 ja 10 vertaamalla näkee puudiagrammiin tapahtuneet muutokset. Harmittavasti, visualisointien painallusten luomaa suodattamista ei saa toimimaan raportin sivujen välille. Jos patenttiluokituksilla haluaa suodattaa, on käytettävä raportin Patent Filters -sivua, jonka suodattimet koskevat kaikkia sivuja.



### 6.3 Patents Table -sivu

Lens Link	Patent Title	Group	Abstract
<a href="https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167">https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167</a>	- Heat-curable bio-based casting composition molding produced therefrom and method for producing such a molding		Disclosed is a heat-curable bio-based casting composition, comprising biomonomers of vegetable or animal origin, (b) one or more polymer: from recycled material or of vegetable or animal origin, (c) inorganic f and methacrylic biomonomer(s) is 10-40% by weight, the proportion c is 44-89% by weight.
<a href="https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167">https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167</a>	- Heat-curable bio-based casting composition molding produced therefrom and method for producing such a molding	Compositions of homopolymers or copolymers of compounds having one or more unsaturated aliphatic radicals, each having only one carbon-to-carbon double bond, and only one being terminated by only one carboxyl radical, or of salts, anhydrides, esters, amides, imides or nitriles thereof; Compositions of derivatives of such polymers	Disclosed is a heat-curable bio-based casting composition, comprising biomonomers of vegetable or animal origin, (b) one or more polymer: from recycled material or of vegetable or animal origin, (c) inorganic f and methacrylic biomonomer(s) is 10-40% by weight, the proportion c is 44-89% by weight.
<a href="https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167">https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167</a>	- Heat-curable bio-based casting composition molding produced therefrom and method for producing such a molding	Compositions of mortars, concrete or artificial stone, containing only organic binders {, e.g. polymer or resin concrete (mechanical aspects of moulding polymer or resin concrete B29C67/242)}	Disclosed is a heat-curable bio-based casting composition, comprising biomonomers of vegetable or animal origin, (b) one or more polymer: from recycled material or of vegetable or animal origin, (c) inorganic f and methacrylic biomonomer(s) is 10-40% by weight, the proportion c is 44-89% by weight.
<a href="https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167">https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167</a>	- Heat-curable bio-based casting composition molding produced therefrom and method for producing such a molding	Copolymers of compounds having one or more unsaturated aliphatic radicals, each having only one carbon-to-carbon double bond, and only one being terminated by only one carboxyl radical or a salt, anhydride ester, amide, imide or nitrile thereof	Disclosed is a heat-curable bio-based casting composition, comprising biomonomers of vegetable or animal origin, (b) one or more polymer: from recycled material or of vegetable or animal origin, (c) inorganic f and methacrylic biomonomer(s) is 10-40% by weight, the proportion c is 44-89% by weight.
<a href="https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167">https://www.lens.org/lens/search/patent/list?q=lens_id:%22082-942-661-517-167</a>	- Heat-curable bio-based casting composition molding produced therefrom and method for producing such a molding	Domestic plumbing installations for fresh water or waste water; Sinks {(work tops A47B77/022)}	Disclosed is a heat-curable bio-based casting composition, comprising biomonomers of vegetable or animal origin, (b) one or more polymer: from recycled material or of vegetable or animal origin, (c) inorganic f and methacrylic biomonomer(s) is 10-40% by weight, the proportion c is 44-89% by weight.

Kuva 11. Raportin Patent Table -sivun tietopöytä näkymä.

Taulukossa 11 on neljä kenttää: Lens Link, Patent Title, Group ja Abstract.

Ennen suodattamista taulukko sisältää jokaisen patentin tiedot. Määrällisesti se on aivan liian monta riviä tietoa, jotta niitä voisi tutkia näin yksilökohtaisin tiedoin. Sen vuoksi taulukon käyttötarkoitus on olla viimeinen askel, kun käyttäjä on rajannut patentit vain heitä kiinnostaviin patenttiluokituksiin. Sen jälkeen jäljelle jäävien patenttien nimet, luokitus ja abstract antavat vihjeitä patentin koko sisällöstä, jonka voi hakea avaamalla linkki valitun patentin verkkosivunäkymään.

Lens Link -kentän luominen vaati muutaman lisäoperaation. Hakeutumalla lens.org-tietokanta -verkkosivuille, sekä suorittamalla haun käyttäen yhden patentin lens\_id -kenttää, pystyi hankkimaan url-osoitteen haulle. Tämän jälkeen käyttämällä Power Querya, oli mahdollista kopioida lens\_id -sarake, lisätä linkin alkuosan ennen lens\_id -arvoa sekä linkin loppuosan sen jälkeen. Esimerkki suoritetusta operaatiosta: [https://www.lens.org/lens/search/patent/list?q=lens\\_id:%22\[lens\\_id\]%22](https://www.lens.org/lens/search/patent/list?q=lens_id:%22[lens_id]%22).

## 6.4 Patent Filters -sivu

### Patent Classification Filters

- ^  CHEMISTRY; METALLURGY
  - ∨  ANIMAL OR VEGETABLE OILS, FATS, FATTY SUBS
  - ∨  BIOCHEMISTRY; BEER; SPIRITS; WINE; VINEGAR
  - ∧  CEMENTS; CONCRETE; ARTIFICIAL STONE; CERAMICS
  - ∨  LIME, MAGNESIA; SLAG; CEMENTS; COMPOUNDS
  - ∨  COATING METALLIC MATERIAL; COATING MATERIALS
  - ∨  DYES; PAINTS; POLISHES; NATURAL RESINS; ADHESIVES
  - ∨  ELECTROLYTIC OR ELECTROPHORETIC PROCESSES
  - ∨  FERTILISERS; MANUFACTURE THEREOF
  - ∨  GLASS; MINERAL OR SLAG WOOL

Kuva 12. Patenttien rajaus patenttiluokitusten perusteella.

Kuvassa 12 näkyy suodatin, jossa on listattu patenttiluokitukset ja niiden sisältämät alaluokitukset. Vasemmalla olevasta nuolesta voi sulkea tai avata alaluokituksia, sekä sen viereisestä tick-box valinnasta voi rajata näkymään sen patenttiluokituksen patentit, jolloin ne, joita ei olla valittu, suodattuvat raportin visualisoinneista pois.

### Year

2013 2022

### Quarter

- Neljännes 1
- Neljännes 2
- Neljännes 3
- Neljännes 4

Kuva 13. Patenttien suodattaminen vuoden ja vuosineljännesten perusteella.

## Country

- AP
- AR
- AT
- AU
- BE
- BR
- CA
- CL
- CN
- 

Kuva 14. Patenttien suodattaminen maiden perusteella.



Kuva 15. Kaikkien suodattimien pois päältä asettamisen -painike.

Kuvassa 13 näkyy suodattimet koskien aikajanaa. Jatkokehityksessä voi tarpeen mukaan lisätä myös kuukaudet suodattimiin, mutta tämän projektin rajoituksessa se ei ollut tarpeellista. Kuvassa 14 näkyy maiden suodatin. Harmittavasti, maat ovat maakoodeina, eivätkä kokonaisina maiden niminä. Tämä varmasti hämmentää loppukäyttäjää. Jatkokehityksessä ongelma voitaisiin ratkaista uudella dimensiopöydällä, joka sisältää sarakkeen maakoodille sekä sarakkeen sitä vastaavalle maan nimelle. Kaikki valitut suodattimet saa helposti pois päältä kuvassa 13 esiintyvällä Clear All Filters -painikkeella.

## 7 TUOTOKSET

Tämän opinnäytetyön tuotoksena on syntynyt Power BI -raportti, joka visualisoi patenttidataa ja mahdollistaa niiden seulomisen vain kaikkein kiinnostavimpiin patentteihin. Tätä tuotosta voidaan hyödyntää mihin tahansa Python-hakutyökalun tuomaan patenttidataan, ei ainoastaan toimeksiantajaa kiinnostavaa ekologiseen betoniin rajattuun dataan.

Toimeksiantaja Reseptirobotti-hankkeen antamat patenttihaun avainsanat löytyvät luvusta 4.3. Tässä luvussa esitellään raportissa ilmenevää tietoa keskityen luvussa 4.2. eriteltyjen tutkimuskysymysten vastaamiseen. Tavoitteena on kiteyttää havainnot niin, että toimeksiantaja ja tätä opinnäytetyötä hyödyntävät tahot voivat helposti löytää heitä kiinnostavat työn tuotokset.

## 7.1 Tuotokset ja niiden tulkinta Reseptirobotti-hankkeelle

Taulukko 2. Patenttien lukumäärät ja niiden julkaisumaa jakaumat.

Maa / organisaatio	Julkaisujen lukumäärä	% kaikista julkaisuista
Kaikki patentit	2867	100%
Kiina (CN)	1189	41%
Korea (KR)	488	17%
Yhdysvallat (US)	322	11%
World Intellectual Property Organization WIPO (WO)	154	5%
European Patent Organization (EP)	128	4%

Tutkimuskysymyksiin ”Montako patenttia hakusanoilla löytyi?” ja ”Mistä maista tai organisaatioista on tullut eniten patenttijulkaisuja?” vastaukset löytyvät taulukosta 2. Patenttidadassa on 2867 patenttia. Niiden viisi suurinta julkaisutahoa ovat Kiina, Korea, Yhdysvallat, World Intellectual Property Organization WIPO (suomeksi Maailman henkisen omaisuuden järjestö), sekä European Patent Organization (suomeksi Euroopan patentti organisaatio).

Rajaus on määritetty viiteen suurimpaan tahoon, sillä niiden jälkeen julkaisumäärät alkoivat olla vähäisiä (3 % tai alle patenttijulkaisujen kokonaismäärästä). Listatuissa viidessä korostuu kolme eniten patenttijulkaisuja tekevää maata, joka kenties antaa perspektiiviä sille, missä teknologista innovaatiota

tehdään paljon. Ymmärrettävästi suuret joko maailmaa tai aluetta koskevat organisaatiot ovat myös julkaisseet paljon patenteja.

Taulukko 3. Patenttiluokitukset ja niihin kuuluvien patenttien määrät.

<b>Luokitus</b>	<b>Patenttiluokitusten määrä</b>
Section	9
Class	77
Subclass	187
Group	675
Subgroup	2503

Taulukko 4. Esimerkkejä patenttiluokitusten sisällöistä.

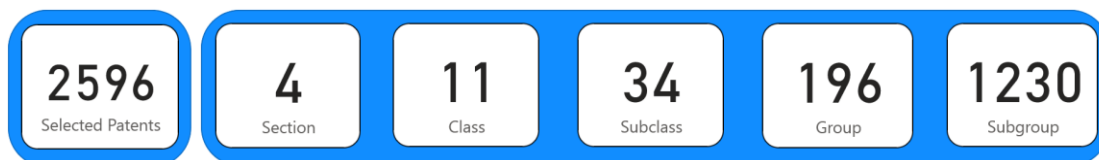
<b>Luokitus</b>	<b>Esimerkkejä patenttiluokituksista</b>
Section	Chemistry; metallurgy, Electricity, Fixed constructions, Human necessities, Mechanical engineering, Performing operations; transporting, Physics, Textiles; Paper...
Class	Cements, Inorganic Chemistry, Agriculture, Natural or man-made threads or fibres, Paper-making; production of cellulose, Building, Construction of roads, railways or bridges...
Subclass	Structural elements; building materials, Lime & Magnesia; slag elements, Compounds of metals..., Compositions of macromolecular compounds, Organic fertilizers..., Finishing work on buildings...
Group	Use of agglomerated or waste materials or refuse as fillers for mortars..., Compositions of mortar, concrete or artificial stone..., Mortars, concretes or artificial stone or mixtures to prepare them...
Subgroup	Waste materials; refuse, Other specific industry waste materials..., Waste from the purification of bauxite...,

Wet materials, e.g. slurries, Flue dust, Slags, Dry materials...

Patenttidatan patenttiluokituksiin jakautuminen näkyy taulukossa 3. Sillä patenttiluokitusten tuloksia on niin monta, tässä tutkimuksessa ei olla tarkemmin eritelty mihin luokituksiin patentit kuuluvat. Kuvassa 4 esitellään esimerkkejä siitä, millaisia patenttiluokituksia patenttidataan kuuluu.

Tärkeä käyttötarkoitus tämän tutkimuksen tuotokselle on raportin kyky suodattaa suurta patenttimäärää pienemmäksi, minkä jälkeen jäljelle jääneitä erityisen kiinnostavia patenteja voi tutkia tarkemmin yksitellen. 4.3. Toimeksiantajan patenttirajaus -luvussa määriteltyä hakusanalista hyödyntäen tutkimuksessa on suoritettu esimerkki patenttien rajauksesta vain oleellisiin patenttiluokituksiin. Samalla, tässä vastataan viimeiseen tutkimuskysymykseen ”Patenttien suodattamisen jälkeen, paljonko aiemmin eriteltyjen patenttijakaumien määrät ovat laskeneet?”

Patenttidatan seassa on useita patenttiluokkia, jotka eivät luultavimmin sisällä toimeksiantajaa kiinnostavia patenteja. Näihin kuuluu mm. metallien- ja jäteveden käsittelyyn liittyvät prosessit, sähköteollisuus, informaatioteknologian käytäntö -patentit, lääketiede ja monia muita. Koska patenttidata on noudettu käyttämällä hakusanoja, ymmärrettävästi mukaan tarttuu kaikenlaisia ei-toivottuja tuloksia.



Kuva 15. Patenttien kokonaismäärä sekä patenttiluokituksiin jakautuminen suodattimien jälkeen.

Taulukko 5. Patenttiluokitusten määrän muutos suodatuksen jälkeen.

<b>Luokitus</b>	<b>Patenttiluokitusten määrä</b>	<b>Suodatettujen patenttiluokitusten määrä</b>	<b>% vähennys kokonaismäärästä</b>
Section	9	4	~55%
Class	77	11	~85%
Subclass	187	34	~81%
Group	675	196	~71%
Subgroup	2503	1230	~51%

Kuvassa 15 näkyy raportin patenttiluokitusten määrien muutos suodattimien aktivoimisen jälkeen. Muutokset ovat kiteytetty taulukossa 5, jossa näkyy patenttiluokitusten määrät, sekä kuinka monta prosenttia vähemmän niitä on suodatusten jälkeen. Valitut suodattimet on valittu arvioiden patenttiluokituksen olennaisuutta ekologiseen betoniin. Suodatuksen jälkeen valtava osa luokituksista suodattui pois. Suuresta patenttiluokkamäärän vähentymisestä huolimatta, kokonaispatenttimäärä väheni 2867 patentista 2596 patenttiin, vain noin 10 %:n vähennys.

Tästä voi päätellä, että valtaosa patenttihaun tuloksista on samoissa patenttikategorioissa. Suodatetut patenttikategoriat olivat epäolennaisia, mutta sisälsivät suhteellisen vähän patenteja. Noin 2,5 tuhatta patenttia on siltikin niin suuri määrä, että niitä on hankala seuloa yksilötasolla. Toimeksiantaja voi halutessaan rajata patentit mm. vuoden ja julkaisumaan avulla, jolloin patenttien määrää voisi vielä kaventaa. Vaihtoehtoisesti, kaikkien kiinnostavien patenttiluokitusten sijasta voisi valita yhden kerrallaan, jolloin Patents Table -sivulla olevien tulosten määrä voi olla toivottavan kapea.

## **7.2 Ehdotukset jatkokehitykselle ja oma pohdinta**

Luomaani Power BI -raporttia voisi kehittää pääosin palvelimella tapahtuvaan osaan. Haetut patentit voisivat mennä tietokantaan, josta niitä voisi visualisoida Power BI:n avulla automaattisesti. Tämän tutkimustyön päätteessä uusi patenttihaku aina syrjäyttää vanhat patenttiedot, sen sijaan että kaikki tallen-

tuisivat. Kenties tulevaisuudessa Power BI -raportissa voisi valita mitä patenttihakua halutaan tutkia, tai kenties kaikkien hakujen patenttidata näkyy raportissa, mutta sitä rajataan jatkokehitetyillä suodattimilla.

Patenttihakua voitaisiin suorittaa Power BI:ssa. Kuvittelen uuden osion raporttiin, joka on kuin hakulomake, johon syötetään hakusanat, patenteja poissulkevat hakutermit ja halutut patenttiluokat. Tämän jälkeen palvelimella toimiva hakutyökalu saa käskyn suorittaa haun, sekä päivittää palvelimella olevaan tietokantaan uudet tiedot, joiden saapumisesta Power BI ilmoittaa ”Hakunne on valmis ja tässä on raportti patenteista.”

Patenttien seulomisen mahdollistamisen tärkeys korostuu ihan missä tahansa tutkimus- ja kehityshankkeessa. Tämän vuoksi, koen tärkeäksi jatkaa tämän projektin kehittämistä jopa Reseptirobotti-hankkeen ulkopuolella. Liian usein hankkeet loppuvat ja vaikka jatkosuunnitelma olisikin tehty, sen toteutus uupuu, jolloin osa hankkeen arvokkaista tuotoksista jää käyttämättä ja näin valtava määrä potentiaalia menee hukkaan.

Tämän tutkimuksen aikana olen oppinut valtavasti. Vaikka omasinkin ennalta kymmenen kuukautta kokemusta tutkimus- ja kehittämistoiminnan parissa teollisten julkaisujen projektinhallinnasta ja kirjoitusprosessista, tämä opinnäytetyö myöhästyi suunnitellusta aikataulusta. Samalla työn osalliset tahot kehittivät ja työn merkitys selkeentyi.

Minun taitoni Power BI -ohjelmiston kanssa kehittyivät valtavasti. Olen siitä ylpeä, sekä uskon sen palvelevan minua paljon työelämässä. Projektinhallintataitoni kehittyivät, etenkin erilaisista jumiutumisen tiloista etenemisen taitojen suhteen. On oma taitonsa osata löytää ratkaisuja ongelmiin, joita ei välttämättä edes täysin ymmärrä. Joskus se vaatii avun pyytämistä, joka ei ole aina helppoa.



## LÄHTEET

About The Lens. 2024. Lens Org. WWW-dokumentti. Saatavissa: <https://about.lens.org/> [viitattu 12.3.2024].

Basic concepts for designers in the Power BI service. 2023a. Microsoft. WWW-dokumentti. Saatavissa: <https://learn.microsoft.com/en-us/power-bi/fundamentals/service-basic-concepts> [viitattu 25.3.2024]

Business intelligence: A complete overview. 2024. Tableau. WWW-dokumentti. Saatavissa: <https://www.tableau.com/learn/articles/business-intelligence#what-is> [viitattu 27.3.2024]

DataLab. 2024. DataLab kuvaus. WWW-dokumentti. Saatavissa: <https://blogit.xamk.fi/datalab/tietoja/> [viitattu 12.3.2024].

EURA2021. 2014. Hanketietopalvelu. Saatavissa: <https://eura2021.fi/hanketietopalvelu?koodi=A80934> [viitattu 22.3.2024]

Get started with Power BI desktop. 2023b. Microsoft. WWW-dokumentti. Saatavissa: <https://learn.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started> [viitattu 27.3.2024]

How to Clean Data in Power BI. 2024. Process.st. WWW-dokumentti. Saatavissa: <https://www.process.st/how-to/clean-data-in-power-bi/> [viitattu 1.4.2024]

Kosonen, M. 2019. Tiedolla johtamisen käsikirja. Kaakkois-Suomen ammattikorkeakoulu. Saatavissa: <https://www.theseus.fi/bitstream/handle/10024/227003/URNISBN9789523441835.pdf?sequence=2&isAllowed=y> [viitattu 26.3.2024]

Neuvonen, A. 2010. Patentti ja markkinatiedon vaikuttavuus pk-yritysten innovaatiotoiminnassa. Artikkel. WWW-dokumentti. Saatavissa: <https://journal.fi/inf/article/view/3564/3315> [viitattu 25.10.2023]

Risch, J., Garda, S., Krestel, R. 2020. ResearchGate. *Hierarchical Document Classification as a Sequence Generation Task*, 2.

Tips for designing a great Power BI dashboard. 2023c. Microsoft. WWW-dokumentti. Saatavissa: <https://learn.microsoft.com/en-us/power-bi/create-reports/service-dashboards-design-tips> [viitattu 25.3.2024]

Tutkimusprosessi. 2024. Tietoarkisto. WWW-dokumentti. Saatavissa: <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/tutkimus/prosessi/> [viitattu 30.3.2024]

Understand star schema and the importance for Power BI. 2023d. Microsoft. WWW-dokumentti. Saatavissa: <https://learn.microsoft.com/en-us/power-bi/guidance/star-schema> [viitattu 1.4.2024]

Viite-ehyden säilyttäminen tietokantamallikaavioissa. 2024. Microsoft. WWW-dokumentti. Saatavissa: <https://support.microsoft.com/fi-fi/office/viite-ehyden-s%C3%A4ilytt%C3%A4minen-tietokantamallikaavioissa-80f60e10-1238-48f7-ab59-2bd31b2f047a> [viitattu 1.4.2024]

Vilkka, H. & Airaksinen, T. 2003. Toiminnallinen opinnäytetyö. Helsinki: Tammi.

What is Power BI? 2024. Microsoft. WWW-dokumentti. Saatavissa: <https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview> [viitattu 27.3.2024]