Rijksdienst voor Ondernemend Nederland

Inzet op sleuteltechnologieën vanuit de WBSO

Resultaten van een haalbaarheidsstudie op basis van text mining

In opdracht van het ministerie van Economische Zaken en Klimaat

>> Duurzaam, Agrarisch, Innovatief en Internationaal Ondernemen

Rijksdienst voor Ondernemend Nederland

Inzet op sleuteltechnologieën vanuit de WBSO

Resultaten van een haalbaarheidsstudie op basis van text mining

Edwin Horlings, Koen Septer, Gerard Schut, Pieter de Bruijn

september 2019

Dit rapport is tot stand gekomen in het kader van het BAT-lab, het beleidsanalyse-laboratorium van het Directoraat-Generaal voor Bedrijfsleven en Innovatie van het Ministerie van Economische Zaken en Klimaat. In dit kader wordt nauw samengewerkt door het Beleidsanalyseteam (BAT) van DG B&I, het Centraal Bureau voor de Statistiek (CBS) en de Rijksdienst voor Ondernemend Nederland (RVO.nl). Centraal in deze samenwerking staan beleidsgedreven analyses op basis van microdata.

Colofon

Projectnaam	Text mining sleuteltechnologieën WBSO
Opdracht en -nummer	Deelopdracht Beleidsanalyse (EZK/B&I 105751)
Versienummer	Definitief september 2019

Auteur	Edwin Horlings, Koen Septer, Gerard Schut, Pieter de Bruijn
Contactpersonen	Pieter de Bruijn (projectleider) pieter.debruijn@rvo.nl 06 2960 3575 Edwin Horlings (onderzoek en hoofdauteur) e.horlings@cbs.nl 070 337 5863 Koen Septer (analist WBSO) koen.septer@rvo.nl 088 042 5104 Gerard Schut (analist WBSO) gerard.schut@rvo.nl 088 042 3276

Deze publicatie is tot stand gekomen in samenwerking met het Centraal Bureau voor de Statistiek en Elsevier Analytical Services in opdracht van het Ministerie van Economische Zaken en Klimaat, Directoraat-Generaal Bedrijfsleven en Innovatie. Naast de hierboven genoemde personen hebben aan deze studie meegewerkt: Marten Kamphorst en Ali Hürriyetoglu van het CBS en Gerrit Jan Bolks, Herman Brouwer, René Mostert, Jurgen Smeenk, Wietze Smit en Bulent Yankaya, de WBSO-adviseurs van RVO.nl. Daarnaast is dankbaar gebruik gemaakt van de expertise van deelnemers aan een expertsessie gehouden op 25 februari 2019 bij RVO.nl in Utrecht.

Rijksdienst voor Ondernemend Nederland
Hanzelaan 310 | 8017 JK Zwolle
Postbus 10073 | 8000 GB Zwolle

De Rijksdienst voor Ondernemend Nederland (RVO.nl) is een agentschap van het Ministerie van Economische Zaken en Klimaat. RVO.nl voert beleid uit voor verschillende ministeries en decentrale overheden als het gaat om duurzaam, agrarisch, internationaal en innovatief ondernemen. RVO.nl is hèt aanspreekpunt voor bedrijven, kennisinstellingen en overheden als het gaat om informatie en advies, financiering, netwerken en wet- en regelgeving.

Inhoud

1 Inleiding

1.1 Aanleiding

Het kabinet zet in op een missiegedreven aanpak in het innovatiebeleid vanuit de overtuiging dat maatschappelijke uitdagingen belangrijke aanjagers zijn voor ons toekomstig verdienvermogen. Technologische doorbraken zijn hierbij onontbeerlijk. Sleuteltechnologieën als fotonica, ICT en kunstmatige intelligentie, nano-, quantum- en biotechnologie zullen de manier waarop we leven, leren, innoveren, werken en produceren ingrijpend veranderen.

"Een sleuteltechnologie is een technologie die gekenmerkt wordt door een breed toepassingsgebied of bereik in innovaties en/of sectoren. Sleuteltechnologieën zijn essentieel bij het oplossen van maatschappelijke uitdagingen en/of leveren een grote potentiële bijdrage aan de economie, door het ontstaan van nieuwe bedrijvigheid en nieuwe markten, het vergroten van de concurrentiekracht, en het versterken van de banengroei. Sleuteltechnologieën maken baanbrekende proces-, product- en/of diensteninnovaties mogelijk. Sleuteltechnologieën zijn relevant voor de wetenschap, maatschappij en de markt." (TNO/NWO, Memo Sleuteltechnologieën, 2017).

De inzet op sleuteltechnologieën en maatschappelijke uitdagingen wordt, in nauwe samenspraak met topteams en departementen, uitgewerkt in de kennis- en innovatieagenda's van de topteams. Deze zijn op hun beurt richtinggevend voor de inzet van investeringen op de topsectoren, door stakeholders als NWO, ZonMW en de TO2-instellingen en het Ministerie van Economische Zaken en Klimaat (EZK) (via bijvoorbeeld PPS-toeslag en MIT). Deze sleuteltechnologieën vormen belangrijke bouwstenen voor de nieuwe aanpak van het innovatiebeleid.

Het Ministerie van EZK wenst inzicht in de mate waarin bestaande innovatieregelingen bijdragen aan sleuteltechnologieën. Zo is door RVO.nl een eerste exercitie uitgevoerd naar de inzet op sleuteltechnologieën vanuit inzetprojecten binnen de PPS-toeslag – waarmee het Ministerie van EZK een impuls biedt aan publiek-private samenwerking in R&D. Naast deze eerste exercitie op het vlak van PPS wenst het ministerie inzicht in de bijdrage van de WBSO aan sleuteltechnologieën; in de eerste plaats omdat de WBSO een generieke regeling betreft die voor alle bedrijven geldt en derhalve een indicatie vormt voor de algehele inspanningen van het Nederlands bedrijfsleven en in de tweede plaats omdat de WBSO – in termen van beleidsbudget – veruit de meest omvangrijke innovatieregeling binnen het EZK-beleidsinstrumentarium is.

De afgelopen jaren is de inzet op thema's[1] en maatschappelijke uitdagingen binnen de WBSO door RVO.nl bepaald op basis van *gewogen trefwoordenanalyse*. Deze methodiek is door RVO.nl ontwikkeld en wordt inmiddels al een aantal jaren gebruikt om WBSO-projecten te classificeren omdat het aantal projecten te groot is om stuk voor stuk te lezen en handmatig in te delen. In deze haalbaarheidsstudie wordt de toepassing van een alternatieve methode onderzocht om te bezien of de inzet op sleuteltechnologieën op een minder arbeidsintensieve wijze en/of meer flexibele wijze in beeld kan worden gebracht en om zo text mining analyses binnen RVO verder te versterken[2].

[1] Thema analyses: Groene Groei, Duurzaamheid (https://www.bedrijvenbeleidinbeeld.nl/bedrijvenbeleid/missiegedreven-topsectoren-en-innovatiebeleid/hoe-staat-nl-ervoor)

[2] Bij de start van deze haalbaarheidsstudie was de wens om op zoek te gaan naar een alternatieve methode ingegeven door het arbeidsintensieve karakter van de methode van gewogen trefwoordenanalyse (met name door de inzet van experts bij het opstellen van trefwoordenlijsten per thema en bij het valideren van de resultaten) en de wens de analyse meer flexibel te maken in relatie tot veranderingen binnen de thema's en classificaties (nieuwe thema's of veranderingen in de scope van bestaande thema's waarvoor bij de methode van gewogen trefwoordenanalyse de analyse in zijn geheel opnieuw moet worden gedaan)

Recent zijn veelbelovende ervaringen opgedaan met het gebruik van text mining door Elsevier in het onderzoek naar de Nederlandse positie op sleuteltechnologieën in de wetenschappelijke literatuur. In opdracht van het ministerie van EZK heeft Elsevier de Nederlandse wetenschappelijke publicaties op het gebied van vijftig sleuteltechnologieën in kaart gebracht. Eerst werden door NWO, TNO en Elsevier relevante trefwoorden en trefwoordentrefzinnen geïdentificeerd. Deze trefwoorden werden vervolgens door deskundigengroepen gevalideerd.[3] Om de trefwoorden te koppelen aan de teksten van wetenschappelijke artikelen heeft Elsevier gebruik gemaakt van een door hen ontwikkelde text mining software tool (open source Python) om zo wetenschappelijke publicaties en citaties in te delen naar sleuteltechnologietechnologie.

1.2 Doel

Het doel van het project is om te verkennen of het mogelijk is om op basis van de methode en tool van Elsevier de inzet op sleuteltechnologieën vanuit de WBSO in kaart te brengen. Deze haalbaarheidsstudie toetst of het mogelijk is om op basis van de Elsevier methodiek en software tool en onderliggende algoritmes bruikbare en betrouwbare resultaten vanuit de database met WBSO-projecten te kunnen produceren met een bescheiden onderzoeksinspanning en met deugdelijke validatie achteraf. Omdat deze studie een haalbaarheidsonderzoek betreft, kunnen inhoudelijke uitkomsten niet op voorhand worden gezien als betrouwbare schattingen of welke andere indicatie dan ook van de beleidsinzet op het totaal van sleuteltechnologieën, hoofdcategorieën van sleuteltechnologieën of specifieke sleuteltechnologieën.

In het project staan dezelfde vijftig sleuteltechnologieën centraal die in het eerdere Elsevier onderzoek op basis van wetenschappelijke publicaties en citaties in beeld zijn gebracht. Deze zijn geclusterd in acht hoofdgroepen en staan beschreven in bijlage 1. Voor de afbakening van deze sleuteltechnologieën is gebruik gemaakt van de trefwoorden zoals die in het eerdere Elsevier onderzoek zijn opgesteld. Deze staan beschreven in bijlage 2.

Het project is nadrukkelijk een *haalbaarheidsstudie*; ten eerste omdat WBSO-projectomschrijvingen verschillen van wetenschappelijke teksten. Het gaat om verschillen in omvang van de teksten, verschillen in taal (Engels versus Nederlands of combinaties van beide) en verschillen in taalgebruik (mate van standaardisatie van centrale begrippen) en ten tweede omdat de analyses vanwege databeveiliging op locatie bij RVO.nl in Zwolle zijn uitgevoerd en het op voorhand niet duidelijk was in hoeverre we de Elsevier tool ook op locatie in Zwolle probleemloos konden toepassen.

1.3 Aanpak en organisatie

Ten behoeve van deze haalbaarheidsstudie zijn onderzoekers van het CBS met text mining expertise gedetacheerd bij RVO.nl in Zwolle, alwaar de analyses zijn verricht op de database van WBSO projecten. De analyses zijn in nauwe samenwerking met analisten van de WBSO gedaan. Hierbij is gebruik gemaakt van de text mining software tool (en onderliggende algoritmes) die door Elsevier beschikbaar is gesteld (het betreft hier een zogenaamde 'light' versie van de tool die ook extern – buiten de servers van Elsevier om – toegepast kan worden). Het uitvoeren van de haalbaarheidsstudie op locatie in Zwolle had als bijkomend voordeel dat tijdens de validatie van de tussenresultaten snel en flexibel geschakeld kon worden met de WBSO-adviseurs (vakinhoudelijke experts op het gebied van de verschillende sleuteltechnologieën), die intensief bij dit onderzoek zijn betrokken.

[3] Dit is gelukt voor 49 van de 50 sleuteltechnologieën; de deskundigen konden het niet eens worden over de trefwoorden voor "Quantumsensoren en -metrologie".

1.4 Leeswijzer

In hoofdstuk 2 wordt de database met WBSO-projecten beschreven. Voor deze studie is de database van 2017 gebruikt waarin 148.892 projecten zijn opgenomen. In hoofdstuk 3 wordt de gewogen trefwoordenmethodiek van RVO.nl beschreven. Dit is de methode die op dit moment wordt gebruikt om te zoeken naar thema's in de WBSO-database. In hoofdstuk 4 wordt beschreven hoe de software tool van Elsevier werkt. De toepassing van de Elsevier software tool op de database met WBSO-projecten en de resultaten die daarmee zijn geproduceerd, worden beschreven in hoofdstuk 5. Hierbij dient in ogenschouw gehouden te worden dat de resultaten niet op voorhand als een indicatie gezien kan worden van de beleidsinzet op sleuteltechnologieën. In hoofdstuk 6 wordt beschreven hoe de WBSO-adviseurs de resultaten hebben gevalideerd en hoe met de informatie die zij hebben aangeleverd een aanvullende analyse is gedaan. De conclusies van de studie zijn te vinden in hoofdstuk 7.

2 De database met WBSO-projecten

Het ministerie van Economische Zaken en Klimaat (EZK) geeft ondernemers de ruimte om te vernieuwen en te groeien, o.a. via de innovatieregeling WBSO. Hiermee kunnen bedrijven een deel van de loonkosten en andere kosten en uitgaven voor R&D-projecten verlagen. De WBSO is al sinds 1994 voor duizenden ondernemers een stimulans om te investeren in R&D. De WBSO vereist dat R&D-projecten voor de start ervan worden aangevraagd. Op die manier wordt vooraf gestimuleerd om meer aan R&D te gaan doen. Technisch experts (WBSO-adviseurs) bij RVO.nl toetsen alle (complete) aanvragen en onderliggende projecten inhoudelijk. Een bedrijf (m.u.v. zelfstandigen) kan maximaal drie keer per jaar een WBSO-aanvraag indienen en elke WBSO-aanvraag kan één of meerdere R&D-projecten bevatten. In 2017 zijn ruim 35.000 WBSO-aanvragen van ruim 21.000 bedrijven (en zelfstandigen) door RVO.nl toegekend. Binnen deze aanvragen zijn 135.900 projecten door de RVO-experts goedgekeurd. De WBSO kent onder R&D-bedrijven in Nederland een hoog doelgroepbereik en is daarmee een zeer interessante bron aan informatie over de private R&D-ontwikkelingen in ons land.

De gebruikte database in dit onderzoek bevat administratieve en inhoudelijke informatie over 148.892 projecten waarvoor in 2017 WBSO is aangevraagd. Deze informatie betreft:

- Administratieve informatie (zoals een aanvraagnummer, een project ID en (afhandelings)status van de aanvraag).
- Primair technologiegebied van de aanvraag (bijv. medische technologie).
- Type project (ontwikkelingsproject of technisch wetenschappelijk onderzoek).
- Zwaartepunt van het project (product, proces of programmatuur).
- Het aantal toegekende uren speur- en ontwikkelingswerk per project.
- Beschrijving van de inhoud van het project. Dit betreft tekstvelden met daarin de titel, een algemene omschrijving en specifieke informatie over, probleemstelling, oplossingsrichting,technische knelpunten, oplossingsrichtingen, toepassing van bestaande en nieuwe methodieken, wijzigingen in de planning (bij doorlopende projecten) en gerichte ICT vragen in het geval van programmatuurprojecten.

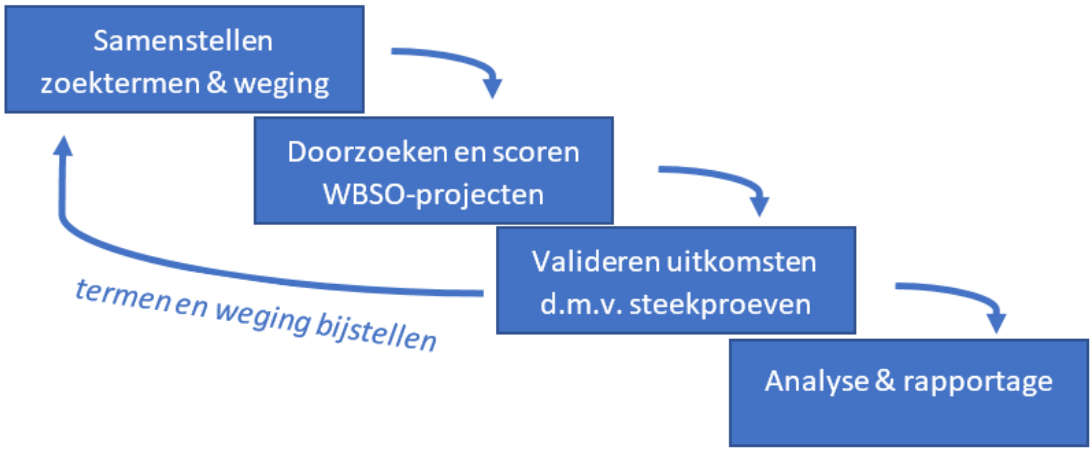
Bij alle aangevraagde projecten kan onderscheid worden gemaakt tussen nieuwe projecten, waarvoor WBSO is aangevraagd en doorlopende projecten waarbij de S&O-werkzaamheden van een specifiek project doorlopen een volgende WBSO-aanvraag. In de gebruikte database kan onderscheid worden gemaakt naar projecten die door RVO zijn goedgekeurd (>0 toegekende S&O-uren) en projecten die zijn afgewezen of ingetrokken door de aanvragen (0 S&O-uren). Voor de analyse met de Elsevier software tool zijn voor project alle 148.892 aangevraagde R&D-projecten in 2017 alle tekstvelden die de inhoud van het project beschrijven samengenomen en geanalyseerd door de tool.

Het bestand met WBSO-projecten is in een SAS-output-file (.sas7bdat) beschikbaar gesteld voor dit onderzoek. Dit bestand is vervolgens in "R" (open source software) ingelezen en vandaar naar een tekstbestand geëxporteerd. Het tekstbestand is vervolgens geschoond. Bedrijfsidentificerende gegevens zoals naam, adres en KVK-inschrijvingsnummer zijn verwijderd. De teksten die aanvragers hebben ingevuld bevatten allerlei tekens en opmaak die moet worden verwijderd. Zo hebben de aanvragers hun teksten in alinea's verdeeld en hebben ze bullets gebruikt. Ook bevat de tekst tabs die in de Elsevier tool worden gebruikt om velden van elkaar te scheiden. De geschoonde versie van de database is zo opgezet dat individuele records (projecten) en individuele velden (de velden met informatie over ieder project) kunnen worden herkend.

3 Gewogen trefwoorden analyse methodiek WBSO

De gewogen trefwoordenmethodiek is door RVO.nl ontwikkeld en wordt al een aantal jaren gebruikt om WBSO-projecten te classificeren. De methodiek bestaat uit vier primaire stappen die in figuur 1 worden weergegeven.

Figuur 1. Schematische weergave van de gewogen trefwoordenmethodiek van RVO.nl



Om verschillende thema's af te bakenen worden voorafgaand aan de analyse voor elk thema een lijst met zoektermen opgesteld, met per zoekterm een wegingsfactor. Dit is een zeer belangrijke stap in het totale proces die in grote mate de kwaliteit van uitkomst bepaald. Deze stap wordt dan ook in samenwerking met een R&D-expert op het gebied van het gevraagde thema uitgevoerd, daardoor kan zoveel mogelijk naar de juiste (technische) vaktermen worden gezocht. Ieder trefwoord krijgt daarnaast een weging op basis van relevantie in relatie tot het gezochte thema. Zoektermen die op zichzelf kenmerkender zijn voor het thema dat wordt gezocht krijgen een hogere waarde dan meer algemene woorden die wellicht wel raakvlak hebben met het gezochte thema, maar mogelijk ook in een andere context een andere betekenis kunnen hebben. In de lijst met zoektermen wordt rekening gehouden met linguïstiek (taal, spellingsvarianten en lemmatisering - het gebruik van woordstammen). In tabel 1 is een (verkorte) zoek termenlijst te zien van het thema Artificial Intelligence waarin de linguïstiek is verwerkt.

Tabel 1. Voorbeeld van gewogen trefwoordenlijst

Zoekterm	Weging
'kunstmatige intelligentie'	10
'artificial intelligence'	10
' AI ' [tussen spaties; bijv. 'za**ai**' uitsluiten]	10
'neuraal netwerk' [enkelvoud]	6
'neurale netwerken' [meervoud]	6
'neural network' ['neural network**s**' is niet nodig; meervoud zit in de enkelvoud stam]	6
'big data'	2
'hersen' [voorbeeld wanneer medisch onderzoek moet worden uitgesloten]	-100

Nadat de gewogen trefwoordenlijst gereed is, wordt deze via SQL-script ingeladen en verwerkt. Het SQL-script scant vervolgens alle WBSO-projectteksten op de gewogen termen en scoort elk project waarin één of meerdere termen worden gevonden op basis van de toegekende weging. Als een project bijvoorbeeld het woord 'kunstmatige

intelligentie' (weging 10) en 'big data' (weging 2) bevat en verder geen andere matches heeft, dan krijgt het een totaal score van 12. Het resultaat van het SQL script is een lijst WBSO-projecten, gecumuleerde score per project en trefwoorden waar een match mee is. Er wordt verondersteld dat projecten met een hoge score ook een hogere waarschijnlijkheid hebben om tot het thema te behoren, er zijn immers meer relevante zoektermen gevonden dan in andere projecten. Met de kennis van de R&D-experts van RVO.nl m.b.t. de zoektermenlijst en het feit dat RVO.nl het SQL-script zelf heeft ontwikkeld en beheert kan flexibel omgegaan worden met verschillende thema's. Tevens is de reproduceerbaarheid van de analysemethode hierdoor gewaarborgd.

Om te kijken of de juiste projecten zijn gevonden en geen projecten zijn gemist (false negatives) en niet teveel projecten zijn gevonden die niet tot het thema behoren (false positives) wordt de hulp van de WBSO-adviseur in dit stadium weer ingeschakeld. Hij of zij neemt systematisch steekproeven van hoge naar lagere projectscores en onderzoekt de gevonden projecten inhoudelijk. Dit is vaak een arbeidsintensief en tijdrovend deel van het totale proces, maar is wel een cruciale stap om iets over de kwaliteit van de uitkomsten te kunnen zeggen. Na inhoudelijke beoordeling wordt een beslisregel vastgesteld en toegepast om te bepalen welke projecten wel en welke niet tot het thema behoren. Deze beslisregel wordt vaak zo bepaald dat er een goede balans is tussen false positives en false negatives. In de praktijk wordt gekozen voor de laagste projectscore waar het aandeel false positives minder dan 1/3 bedraagt. Het oordeel van de RVO-expert is in deze benadering uitermate belangrijk.

In sommige onderzoeken wordt van deze beslisregel afgeweken en bij kleinere thema's worden door de RVO-expert de gevonden projecten integraal inhoudelijk gevalideerd. Indien de RVO-expert van mening is dat er sowieso teveel false positives zijn of juist teveel false negatives zijn over de hele linie van projectscores, dan wordt de zoektermenlijst aangepast (zowel de termen als scores) en wordt het SQL-script opnieuw uitgevoerd. Nadat door middel van een beslisregel is vastgesteld welke WBSO-projecten tot het gevraagde thema behoren wordt de dataset met WBSO-projecten verder verrijkt met aanvullende informatie over de aanvrager, R&D-uren, fiscaal voordeel, enzovoorts en worden met behulp van deze gegevens verdere analyses en rapportages opgesteld. Ook bestaat de mogelijkheid om de analyses met terugwerkende kracht op eerdere WBSO-jaren toe te passen, waarmee de ontwikkeling/trend van een thema in beeld kan worden gebracht. Zodra een thema correct is afgebakend en er een beslisregel is vastgesteld, kan relatief eenvoudig een nieuwe scan en analyse worden gemaakt van een nieuw jaar, omdat de zoektermenlijst gelijk kan blijven en de arbeidsintensieve steekproeven kunnen worden overgeslagen.

4 De software tool van Elsevier

De werking van de software tool en onderliggende algoritmes van Elsevier bepaalt de aard van de resultaten. In dit hoofdstuk wordt beschreven hoe de software tool werkt en wat in dit project is gedaan om de tool op locatie bij RVO.nl in Zwolle werkend te krijgen.

4.1 Installatie van de Elsevier software tool en de zoektermen

Elsevier heeft een software tool (Python) geleverd die:

- werkt vanuit een vooraf opgestelde lijst met zoektermen voor de 50 sleuteltechnologieën;
- in de projectbeschrijvingen zoekt op het voorkomen van de zoektermen; en
- als resultaat teruggeeft de *identifier* (ID) van het project, de ID van de betreffende sleuteltechnologie, de specifieke zoekterm en het aantal matches.

De lijst met sleuteltechnologieën staat in bijlage 1. De lijst met zoektermen per sleuteltechnologie staat in bijlage 2.

De gebruikte Elsevier software tool is een 'light' versie. Deze versie normaliseert enkel- en meervoud (bijv. 'child' en 'children'), werkwoordvervoegingen (bijv. 'clone', 'cloned' en 'cloning'), taal- en spellingsvarianten (bijv. 'gynaecology' en 'gynecology'), en koppeltekens aan het einde van een regel (bijv. 'dehyph- enation' en 'dehyphenation'). Wat de 'light' versie niet doet (en de volledige versie wel) is afkortingen detecteren en uitschrijven (bijv. 'BG' en 'Blood Group'), coördinaties detecteren en uitschrijven (bijv. 'intra- and extramural' en 'intramural' en 'extramural') en termen herschrijven (bijv. 'alpha level' en 'α level'). Wat zeer nadrukkelijk moet worden gezegd, is dat de Elsevier software tool alleen Engelstalig tekst kan normaliseren. Nederlandstalige tekst wordt geanalyseerd zoals het staat geschreven, zonder het toepassen van linguïstiek.

RVO.nl heeft recent voor o.a. de WBSO-evaluatie en dit project een snelle stand-alone desktop computer aangeschaft en deze ook voor dit project beschikbaar gesteld[4]. Deze desktop heeft voldoende snel geheugen en multicore rekenkracht om de grote hoeveelheden data in relatief korte tijd te kunnen verwerken.

De softwaretool die Elsevier had geleverd werkte in eerste instantie niet. Een aantal modules kon niet worden geïnstalleerd. Elsevier heeft de software op een andere manier beschikbaar gemaakt en een versie geleverd die veel sneller werkte.[5]

Wel bleek dat de tool nog niet het volume van de hele database aankon. Om Elsevier te helpen bij het oplossen van het probleem is een tekstbestand toegestuurd met alle openbare samenvattingen van Horizon 2020 projecten in 2017 (18.925 zeer tekstrijke records). Daarmee kon het probleem door Elsevier worden opgelost. Uiteindelijk is het gelukt om de hele database in twee blokken van iets meer dan 70 duizend projectbeschrijvingen te analyseren.

4.2 De zoektermen

Elsevier heeft de zoektermen geleverd die zijn gebruikt om in wetenschappelijke publicaties te zoeken naar sleuteltechnologieën. Deze zoektermen zijn met textmining algoritmes uit de tekst van wetenschappelijke publicaties gehaald en gevalideerd door wetenschappelijke experts.

[4] Zie Dialogic, APE en UNU-MERIT (2018) Evaluatie WBSO 2011-2017, pp. 81-85 voor text mining toepassingen om het belang van programmatuur en digitiale innovatie in de WBSO in kaart te brengen

[5] Via Docker, een programma dat software van derden laat draaien in een virtuele omgeving, zonder dat die software op de betreffende pc geïnstalleerd hoeft te worden.

Het betreft 1.579 zoektermen die zijn gekoppeld aan 50 sleuteltechnologieën (zie bijlage 2 voor een overzicht). Een aantal zoektermen heeft betrekking op meer dan één sleuteltechnologie. Sommige sleuteltechnologieën delen een groot aantal zoektermen. Bijvoorbeeld, de zoektermen "biofabrication" en "bioprinting" verwijzen naar vijf sleuteltechnologieën: "biofabrication", "organ on a chip", "additive manufacturing/3D printing", "bio (related) materials and soft material", en "designer and meta materials". De zoekterm "Lab-on-a-chip" verwijst naar "Micro and nanofluidics", "Nanomedicine", "Organ on a chip", "Biochips and biosensors" en "Microreactors".[6]

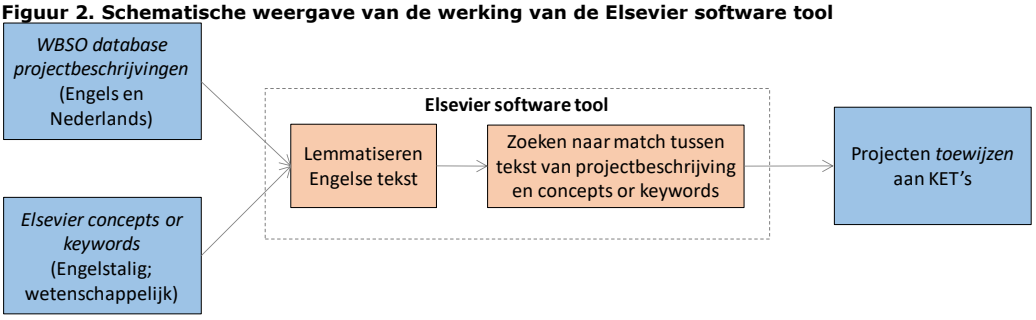
De zoektermen van Elsevier zijn exact gebruikt zoals ze zijn aangeleverd. De lijst is niet uitputtend. Niet iedere relevante variant van iedere zoekterm is gegeven. Zo komen sommige zoektermen alleen in het meervoud voor (bijvoorbeeld, "Cadmium Sulfide Solar Cells" bij Thin films and coatings) en andere zowel in het enkelvoud als het meervoud (bijvoorbeeld "Side Channel Attack" en "Side Channel Attacks" bij Encryption technologies/ digital security). Afkortingen worden niet consequent toegepast. Zo wordt bij Genomics/proteomics/metabolomics/ glycomics/X-omics wel "next generation sequencing" gebruikt maar niet de afkorting NGS; wordt bij Synthetic cell technology de zoekterm "CRISPR technology" wel gebruikt maar "CRISPR" op zichzelf niet, terwijl dit bij Gene editing/precise genetic engineering juist andersom is. Er is vanuit gegaan dat alle keuzes ten aanzien van de zoektermen zijn voortgekomen uit het validatieproces met de experts van NWO en TNO. Daarom is besloten hier niets aan te veranderen.

4.3 Hoe werkt de tool?

De Elsevier software tool is een set van algoritmes (Python) waarmee een tekstbestand kan worden doorzocht op de aanwezigheid van een vooraf bepaalde verzameling zoektermen (een 'vocabulaire' van zoektermen). De zoektermen kunnen eenvoudig zijn (een enkel woord) maar ook complex (combinaties van meerdere woorden). De tool maakt de vocabulaire en het tekstbestand vergelijkbaar en geeft als output de overeenkomsten ('matches') die zijn gevonden.

De tool werkt met tekstbestanden (zogenaamde .tsv of tab-separated values). Daarin staat een project_id gevolgd door een tab met vervolgens alle tekstuele informatie waarin naar KET-zoektermen gezocht gaat worden.

Figuur 2. Schematische weergave van de werking van de Elsevier software tool



De software tool werkt in vijf stappen (figuur 2):

1. De vocabulaire wordt ingelezen. Dit is de lijst met zoektermen die bij iedere sleuteltechnologie horen.
2. Het tekstbestand met projectbeschrijvingen wordt ingelezen.

[6] Het rapport van Elsevier geeft aan dat een perfecte scheiding of afbakening van sleuteltechnologieën niet mogelijk is. Het gevolg is dat de zoektermen van sleuteltechnologieën overlappen. Zie figuur 2 op pagina 5 van Elsevier (2018).

De vocabulaire en het tekstbestand worden door de gebruiker bepaald. Het is in principe mogelijk om de WBSO-projectbeschrijvingen te doorzoeken met een andere lijst zoektermen. Dat is in deze studie niet gedaan.

3. De inhoud van de vocabulaire (lijst met zoektermen) wordt gelemmatiseerd en vervolgens genormaliseerd. Lemmatiseren houdt in dat woorden worden herleid tot het lemma in het woordenboek waarvan zij zijn afgeleid. Bijvoorbeeld, "models" wordt herleid tot "model", "analysed" wordt herleid tot "analyse", enzovoorts. Normaliseren is een proces waarbij verschillende schrijf- en spellingsvarianten van een woord naar een eenduidige variant worden omgezet. Bijvoorbeeld, "modelling" wordt omgezet naar "modeling", "aeroplane" wordt omgezet naar "airplane", "US$" to "$", enzovoorts.
4. De woorden in het tekstbestand (projectomschrijvingen) worden gelemmatiseerd en vervolgens genormaliseerd.

Na lemmatisering en normalisering zijn de woorden in de vocabulaire en het tekstbestand vergelijkbaar gemaakt. ***Het is belangrijk om te benadrukken dat het lemmatiseren en normaliseren alleen wordt gedaan voor Engelstalige tekst. Lemmatisering en normalisering van Nederlandstalige teksten wordt niet ondersteund.***

5. Vergelijken van de gelemmatiseerde en genormaliseerde varianten van de vocabulaire en de projectbeschrijvingen. Als een match wordt gevonden, schrijft de tool de ID van het project, de ID van de sleuteltechnologie, de specifieke zoekterm en het aantal matches.

4.4 De output

De Elsevier tool produceert een tekstbestand waarin de projecten worden gekoppeld aan een of meerdere sleuteltechnologieën. Tabel 2 geeft een voorbeeld van de output. Bijvoorbeeld, in de beschrijving van project 39xxx30 zijn drie zoektermen gevonden ("laser", "nanomaterial", "Optical imaging") die horen bij vier sleuteltechnologieën: Imaging technologies (10), Photon generation technologies (18), Photonic detection (19) en Nanomaterials (44).

Tabel 2. Voorbeeld van de output van de Elsevier tool

TextId	KetId	MatchText	MatchTextCount
39xxx27	18	laser	2
39xxx27	29	CAM	1
39xxx28	18	laser	2
39xxx30	18	laser	2
39xxx30	44	nanomaterial	1
39xxx30	10	Optical imaging	1
39xxx30	19	Optical imaging	1
39xxx40	6	control algorithm	6
39xxx76	18	laser	2
39xxx76	29	CAM	1
39xxx77	18	laser	2
39xxx79	18	laser	2
39xxx79	44	nanomaterial	1
39xxx79	10	Optical imaging	1
39xxx79	19	Optical imaging	1

4.5 Toewijzing van projecten aan sleuteltechnologieën

Met de output van de Elsevier software tool kan een project worden toegeschreven aan een of meerdere sleuteltechnologieën. De tool constateert slechts dat er een match is tussen een project en een sleuteltechnologie op een bepaalde zoekterm en geeft het

aantal keer dat de betreffende term in de tekst is gevonden. De gebruiker moet vervolgens zelf de grenswaarde bepalen (van een ook door de gebruiker te bepalen indicator op basis van het aantal verschillende trefwoorden binnen een sleuteltechnologie dat gevonden is in combinatie met het aantal keer dat deze trefwoorden zijn gevonden) om projecten aan sleuteltechnologieën toe te wijzen. In dit opzicht werkt de Elsevier software tool net als de gewogen trefwoorden analyse methodiek. De toewijzingsmethode kan grote impact hebben op de uiteindelijke resultaten.

5 Resultaten

In dit hoofdstuk presenteren we de resultaten van de toepassing van de software tool en de zoektermen van Elsevier op de database met alle 148.892 WBSO-projecten uit 2017. Voor de analyse met de Elsevier tool zijn alle tekstvelden die voor een WBSO-project beschikbaar zijn (o.a. titel, algemene omschrijving, technische knelpunten, nieuwheid, etcetera) samengenomen.

5.1 De taal van projectbeschrijvingen

Taal is een cruciaal aspect van de analyse. Vooraf was de verwachting dat een-op-de-vijf WBSO-projectbeschrijvingen Engelstalig zou zijn. De zoektermenzoektermen van Elsevier zijn Engelstalig.

Elsevier heeft in zijn softwaretool taalherkenning ingebouwd (door het aanroepen van beschikbare Python toolkits). Daarmee wordt herkend of een projectbeschrijving is geschreven in het Nederlands of het Engels. Met dezelfde Python module (*langdetect*) is door ons een Python script geschreven die de taal van alle projectbeschrijvingen herkent en in een tabel opslaat. Met die tabel kunnen we zelf selecties maken van Nederlandstalige en Engelstalige projecten. Het komt regelmatig voor dat projectbeschrijvingen zowel Engelse als Nederlandse tekst bevatten. De specifieke module herkent de dominante taal.

5.2 Toepassing op Engelstalige projectbeschrijvingen

De resultaten van de toepassing van de Elsevier software tool op de Engelstalige projectbeschrijvingen zijn in het kort als volgt:

- Van de 148.892 WBSO-projecten hebben 5.896 (4 procent) een Engelstalige beschrijving.
- Van deze 5.896 projecten is voor de helft (2.949 projecten of 50 procent) door de Elsevier tool een match gevonden met een sleuteltechnologie.
- Van de 50 sleuteltechnologieën zijn er 49 aangetroffen. De enige die niet werd gevonden is "Quantum sensors and metrology". Dit is de enige sleuteltechnologie waar de experts het niet eens konden worden over de zoektermen. De zoektermenlijst voor deze specifieke sleuteltechnologie bestaat uit slechts drie zoektermen ("Nanomechanical quantum systems", "quantum metrology" en "quantum sensor").
- Van de 1.579 zoektermenzijn er in de Engelstalige projectbeschrijvingen 362 (23 procent) gevonden en 1.217 (77 procent) niet.

5.3 Toepassing op Nederlandstalige projectbeschrijvingen

De resultaten van de toepassing van de Elsevier software tool op de Nederlandstalige projectbeschrijvingen zijn in het kort als volgt:

- Van de 148.892 WBSO-projecten hebben 142.997 (96 procent) een Nederlandstalige beschrijving.
- Van deze projecten is 15 procent (21.239 projecten) door de Elsevier tool een match gevonden met een sleuteltechnologie. Het gaat hierbij dus om Engelse zoektermen in een projectomschrijving waarin Nederlands de dominante taal is. Dit percentage is veel lager dan dat voor de Engelstalige projecten. In totaal zijn 121.578 projecten (81 procent van alle WBSO-projecten in de database) niet aan een sleuteltechnologie gekoppeld.
- Van de 50 sleuteltechnologieën zijn er 49 aangetroffen. De enige die niet werd gevonden is "Quantum sensors and metrology". Dit resultaat komt overeen met dat voor de Engelstalige projecten.
- Van de 1.579 Engelstalige zoektermen van Elsevier zijn er in de Nederlandstalige projectbeschrijvingen 406 (26 procent) gevonden en 1.173 (74 procent) niet. Het

aantal gevonden zoektermen is hoger dan in het geval van de Engelstalige projectbeschrijvingen. Dit hangt waarschijnlijk samen met het veel grotere aantal Nederlandse projectbeschrijvingen.

5.4 Toepassing op alle projectbeschrijvingen

Wanneer we de resultaten van de Nederlandstalige en Engelstalige projectbeschrijvingen combineren, vinden we het volgende:

- Op-een-na zijn alle sleuteltechnologieën gevonden.
- Van de 148.892 WBSO-projecten is voor 24.188 projecten (16 procent) een match gevonden met een sleuteltechnologie.
- Van de 1.579 zoektermen van Elsevier zijn er 511 gevonden (32 procent) en 1.068 (68 procent) niet. Het percentage dat wordt gevonden lijkt verband te houden met het aantal woorden in een zoekterm. Tabel 3 laat zien dat het percentage dat wordt gevonden afneemt naarmate de zoekterm bestaat uit meer woorden.

Tabel 3. Relatie tussen het aantal woorden in Elsevier's zoektermenlijstzoektermen en het percentage dat is gevonden

Woorden per zoektermen	Totaal aantal unieke zoektermenlijst	Aantal gevonden unieke zoektermenlijst	Percentage gevonden
1	326	181	56%
2	841	264	31%
3	322	57	18%
4	90	9	10%
Totaal	1579	511	32%

Tabel 4 laat zien dat het percentage van Elsevier's zoektermenlijst dat voor iedere sleuteltechnologie is gevonden sterk uiteenloopt.

- Van de zoektermen voor de sleuteltechnologieën "Advanced materials", "Nanotechnologies", "Photonics and light technologies" en "Quantum technologies" werd minder dan 30 procent gevonden. De zoektermen van "Quantum technologies" zijn relatief het minst gevonden (14 procent).
- Van de zoektermen van "Digital technologies" is meer dan de helft gevonden.
- De gemiddeldes per groep van sleuteltechnologieën vertellen niet het hele verhaal. Binnen iedere groep kunnen de percentages aanzienlijk variëren.
- Er is geen verband tussen het percentage van de lijst met zoektermen dat werd gevonden en het aantal woorden dat ze bevatten. Sleuteltechnologieën waarvan maar een relatief klein deel van de zoektermen werd gevonden, hadden gemiddeld per zoekterm niet meer woorden dan sleuteltechnologieën waarvan een relatief groot deel van de zoektermen werd gevonden.

Een percentage betekent niet noodzakelijk een goede uitkomst in die zin dat deze aanpak accurate resultaten oplevert. Dit verwijst naar de balans tussen *precision* (accurate uitkomsten) en *recall* (een groot aantal gevonden projecten). Een laag percentage betekent niet dat een verzameling zoektermen slechter toepasbaar is. Het betekent alleen dat minder van deze zoektermen in de beschrijving van WBSO-projecten voorkomen. Een enkele term kan een uitstekende indicator zijn voor een sleuteltechnologie. Dit zou kunnen komen doordat de zoektermen zijn opgesteld om te zoeken in de teksten van internationale wetenschappelijke tijdschriften.

Tabel 4. Percentage van Elsevier's zoektermen dat per sleuteltechnologie is gevonden

Gebied	Sleuteltechnologie	Zoektermen	Gevonden	Gem.
Advanced Materials	Bio (related) materials and soft material	22	41%	22%
	Composite and ceramics	39	21%	
	Designer and meta materials	37	11%	
	Energy conversion	32	31%	
	Energy storage materials	55	18%	
	Optical/electronic/magnetic materials (incl 2D and graphene)	65	15%	
	Smart/self healing/self-organizing materials	31	16%	
	Structural materials	25	24%	
	Thin films and coatings	41	22%	
Chemical technologies	(Bio)Process technology including process intensification	33	33%	40%
	Analytic technologies	37	41%	
	Catalysis	40	55%	
	Electrification / Hydrogen technology / power to gas	52	31%	
	Microreactors	36	42%	
	Separation technology	18	39%	
Digital technologies	Artifical intelligence (incl. machine and deep learning)	36	36%	52%
	Big data and data analytics	33	58%	
	Block chain	44	57%	
	Encryption technologies/ digital security	37	43%	
	High Perfomance Computing Grid Computing and Cloud Technologies/Computing	31	68%	
Engineering and fabrication technologies	(Opto)mechatronics	28	25%	37%
	Additive manufacturing/3D printing	32	44%	
	Cyberphysical systems	71	51%	
	High frequency and mixed signal technologies	67	39%	
	Imaging technologies	65	26%	
	Robotics	43	49%	
	Sensors and actuators	22	27%	
Life sciences technologies	Biocatalysis	30	40%	37%
	Biochips and biosensors	29	31%	
	Biofabrication	30	33%	
	Gene editing/precise genetic engineering	25	48%	
	Genomics/proteomics/metabolomics/ glycomics/X-omics	26	50%	
	Industrial biotechnology (white)	19	42%	
	Nanomedicine	31	26%	
	Organ on a chip	26	27%	
	Stem cell technology	17	41%	
	Synthetic cell technology	20	30%	
Nanotechnologies	Bionano	19	16%	27%
	Micro and nanofluidics	16	31%	
	Nanomanufacturing	27	44%	
	Nanomaterials	61	26%	
	Nanoscale devices	11	18%	
	Semiconductor devices	57	26%	
Photonics and light technologies	Integrated photonics	39	31%	25%
	Photon generation technologies	60	22%	
	Photonic detection	41	24%	
	Photovoltaics	93	25%	
Quantum technologies	Quantum communication	13	23%	14%
	Quantum computing	32	19%	
	Quantum sensors and metrology	3	0%	

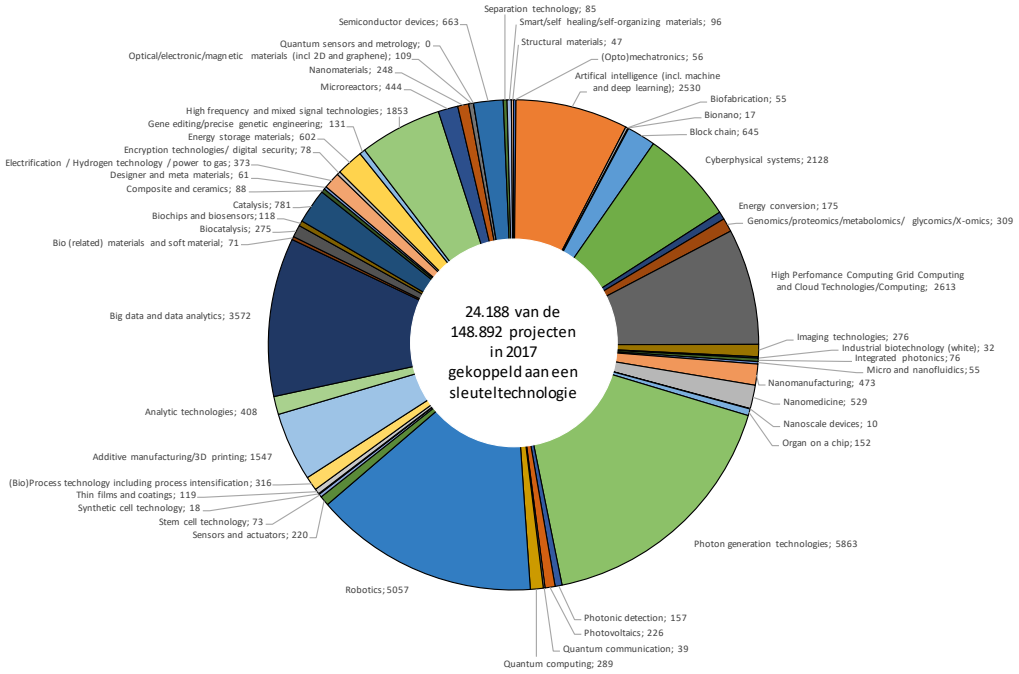
5.5 Toewijzing van gevonden matches aan sleuteltechnologieën

Er zijn verschillende manieren om projecten toe te wijzen aan sleuteltechnologieën op basis van de output van de Elsevier software tool. In deze paragraaf gebruiken we twee methoden om de resultaten toe te wijzen.

De eerste en eenvoudigste methode is om iedere koppeling van een project met een sleuteltechnologie als gelijkwaardig te behandelen, ongeacht het aantal matches met de zoektermen. Een project dat een enkele zoekterm bevat die bij een sleuteltechnologie hoort en een project dat tien zoektermen bevat die meerdere keren in de projectbeschrijving voorkomen worden beide zonder kwalificatie aan de sleuteltechnologie toegewezen.

Het voordeel van de eerste methode is dat alle informatie die is verzameld wordt meegenomen. Alle gevonden projecten worden toegewezen. Het nadeel is dat de precisie van deze methode laag is. Een deel van de koppelingen met sleuteltechnologieën berust op een enkele zoekterm, soms op zeer generieke zoektermen (bijvoorbeeld "detection").

Figuur 3. Alle geclassificeerde Engelstalige en Nederlandstalige projecten die op één of meerdere zoektermen een koppeling hebben met een sleuteltechnologie



De tweede methode is om onderscheid te maken tussen koppelingen van verschillende sterkte. De sterkte van een koppeling is hier afhankelijk van het aantal zoektermen dat is gevonden en het aantal matches per zoektermzoekterm (zie tabel 5 voor een voorbeeld).

- Een project met een *sterke koppeling* met een sleuteltechnologie bevat meerdere zoektermen en heeft op tenminste een zoekterm meerdere matches.
- Een project met een *zwakke koppeling* bevat slechts één enkele zoekterm die eenmaal is gevonden.
- Een project met een *gemiddelde koppeling* is dan logischerwijs een project dat meerdere zoektermen bevat die ieder maar één keer voorkomen.

Tabel 5. Voorbeelden van de resultaten met verschillende sterkten van koppelingen

Voorbeeld van de resultaten van project met een sterke koppeling		*Voorbeeld van de resultaten van project met een gemiddelde koppeling*		*Voorbeeld van de resultaten van project met een zwakke koppeling*	
Monero	1	Ripple	1	File Sharing	1
cryptocurrency	3	litecoin	1		
Ethereum	2	Monero	1		
Ripple	1				
litecoin	1				
blockchain	17				
Bitcoin	4				

Van alle koppelingen tussen een projectbeschrijving en een sleuteltechnologie is maar een klein deel sterk. Het aantal koppelingen is groter dan het aantal projecten, omdat een project aan meerdere sleuteltechnologieën gekoppeld kan worden.

- Van alle 34.158 *koppelingen* tussen WBSO-projecten en sleuteltechnologieën is 8 procent sterk, 38 procent gemiddeld en 54 procent zwak.
- Van alle 24.188 *WBSO-projecten* waarin zoektermenzoektermen van een of meerdere sleuteltechnologieën werden gevonden had 4 procent een sterke *koppeling met een sleuteltechnologie*, 30 procent een gemiddelde *koppeling*, en 66 procent een zwakke *koppeling*.

Als we ons concentreren op de projecten met een sterke *koppeling* met één of meerdere sleuteltechnologieën, dan kunnen van de 148.892 projecten niet meer dan 989 projecten (0,7 procent) worden geclassificeerd.

5.6 Conclusies met betrekking tot de toepassing van de Elsevier software tool

De resultaten van de eerste toepassing van de Elsevier software tool kunnen niet zonder meer worden gebruikt om WBSO-projecten toe te wijzen aan sleuteltechnologieën.
De voornaamste redenen hebben te maken met de taal waarin projectbeschrijvingen zijn geschreven:

- Er is een groot verschil tussen het percentage van de Engelstalige projecten waarin zoektermen voor een sleuteltechnologie werden gevonden (50%) en het percentage van de Nederlandstalige projecten (15%).
- De Elsevier software tool gaat anders om met Engelstalige tekst dan met Nederlandstalige tekst. Engelstalige tekst wordt eerst genormaliseerd alvorens naar zoektermen wordt gezocht. Hierdoor worden zowel de exacte zoektermen als varianten daarvan gevonden. Nederlandstalige tekst wordt niet genormaliseerd. Hierdoor wordt van iedere zoekterm alleen de variant gevonden die wordt gezocht en niet eventuele andere varianten. Het resultaat is daardoor kwalitatief verschillend.
- De zoektermen zijn uitsluitend in het Engels en toegespitst op wetenschappelijke teksten, terwijl 96 procent van de WBSO-projectbeschrijvingen Nederlandstalig is en aanvragen door bedrijven worden gedaan.
- Niet alle zoektermen zijn gelijk. Hoe groter het aantal woorden in een zoekterm, des te lager het percentage van de zoektermen dat in de database wordt gevonden. Met complexe zoektermen (bestaande uit 2, 3 of 4 woorden) kan preciezer worden omschreven welke projecten moeten worden gevonden (hogere *precision*) maar verkleinen de kans dat die projecten worden gevonden (lagere *recall*).

Een ander probleem betreft de toewijzing van projecten aan sleuteltechnologieën op basis van de gevonden overeenkomsten. Net als bij de gewogen trefwoorden analyse methodiek levert de output van de Elsevier software tool alleen informatie over de match tussen een project en de zoektermen behorend bij een sleuteltechnologie. Er is geen harde statistische maatstaf of grenswaarde waarmee kan worden bepaald of een project bij een sleuteltechnologie hoort of niet. Wanneer alleen die projecten worden geselecteerd die een sterke koppeling met een sleuteltechnologie hebben (meerdere zoektermen

gevonden waarvan minimaal een meer dan eens voorkwam in de tekst), dan blijft **minder dan één procent** van de projecten over.

De vraag wanneer een project wel of niet bij een sleuteltechnologie hoort is niet te bepalen zonder hulp van experts. *Precision* en *recall* kunnen niet beoordeeld worden zonder referentiepunt. Daarom is voor een negental sleuteltechnologieën aan WBSO-adviseurs gevraagd om de resultaten van de eerste analyseronde te valideren.

6 Validatie door WBSO-adviseurs

De WBSO-adviseurs zijn experts op het gebied van R&D met betrekking tot specifieke (sleutel)technologieën. Zij kunnen als geen ander aan een projectbeschrijving zien bij welke technologie een project hoort of juist niet hoort. Een aantal experts is gevraagd om de resultaten van deze haalbaarheidsstudie te valideren.

De validatie door WBSO-adviseurs betrof twee rondes. Eerst hebben de adviseurs de resultaten die met de Elsevier software tool zijn geproduceerd beoordeeld. Vervolgens hebben ze een aangepaste lijst met zoektermen beoordeeld. Met die nieuwe zoektermen is vervolgens een vervolganalyse gedaan.

De validatie is uitgevoerd voor negen sleuteltechnologieën:
- Additive manufacting/3D printing (*engineering and fabrication technologies*)
- Artificial intelligence (incl. machine and deep learning) (*digital technologies*)
- Big data and data analytics (*digital technologies*)
- Energy storage materials (*advanced materials*)
- Gene editing/precise genetic engineering (*e*)
- Genomics/proteomics/metabolomics/glycomics/X-omics (*life science technologiese*)
- High frequency and mixed signal technologies (*engineering and fabrication technologies*)
- Imaging technologies (*engineering and fabrication technologies*)
- Thin film and coatings (*advanced materials*)

6.1 Validatie van geclassificeerde projecten

Aan iedere WBSO-adviseur is een verzameling projecten gegeven om te valideren. Iedere verzameling bevatte ten hoogste 150 projecten: maximaal 50 met een sterke koppeling met de sleuteltechnologie, maximaal 50 met een zwakke koppeling en maximaal 50 met een gemiddelde koppeling. In totaal hebben de adviseurs 1.001 projecten beoordeeld.

De experts is gevraagd om twee vragen te beantwoorden:
1. *Zijn de projecten die wij met de methode van Elsevier hebben toegeschreven aan de KET's (Key Enabling Technologies) correct toegeschreven?* Deze vraag is gesteld om het aantal zogeheten "type I fouten" te verkleinen. Dat zijn 'false positive' resultaten, waarbij projecten ten onrechte worden toegeschreven aan een KET.
2. *Hoe herkennen we projecten die bij een KET zouden horen maar nu niet worden gevonden?* Deze vraag is gesteld om het aantal zogeheten "type II fouten" te verkleinen. Dat zijn 'false negative' resultaten, waarbij projecten niet worden toegeschreven aan een KET volgens de text mining algoritmes terwijl dat wel had moeten gebeuren. Dit kan bijvoorbeeld gebeuren als de gebruikte trefwoordenlijsten (ontwikkeld door Elsevier op basis van Engelstalige wetenschappelijke publicaties) belangrijke zoektermen missen die in de WBSO wel veel gebruikt worden. De WBSO-adviseurs is gevraagd om termen en concepten waarmee hun specifieke sleuteltechnologie beter gevonden kan worden, in het bijzonder Nederlandstalige trefwoorden omdat de aanvankelijke lijst van Elsevier daar niet in voorzag.

Vraag 1 heeft betrekking op precision (vinden we *precies* wat we zoeken?). Vraag 2 heeft betrekking op recall (hoe zorgen we dat we *zoveel mogelijk* vinden wat we zoeken?).

Figuren 4 tot en met 7 vatten de resultaten van de validatie samen. Ze laten zien welk percentage van de beoordeelde projecten – in totaal, met een sterke koppeling, met een gemiddelde koppeling en met een zwakke koppeling met de sleuteltechnologie – volgens de WBSO-adviseurs wel of niet aan de betreffende sleuteltechnologie moet worden toegewezen.

Figuur 4 laat zien dat bij vijf sleuteltechnologieën een substantieel deel van de beoordeelde projecten volgens de experts niet correct is toegewezen. Dit betreft:
- Artificial intelligence (incl. machine and deep learning) (62 procent "nee").
- Big data and data analytics (75 procent "nee").
- High frequency and mixed signal technologies (65 procent "nee").
- Additive manufacting/3D printing (45 procent "nee").
- Energy storage materials (67 procent "nee").

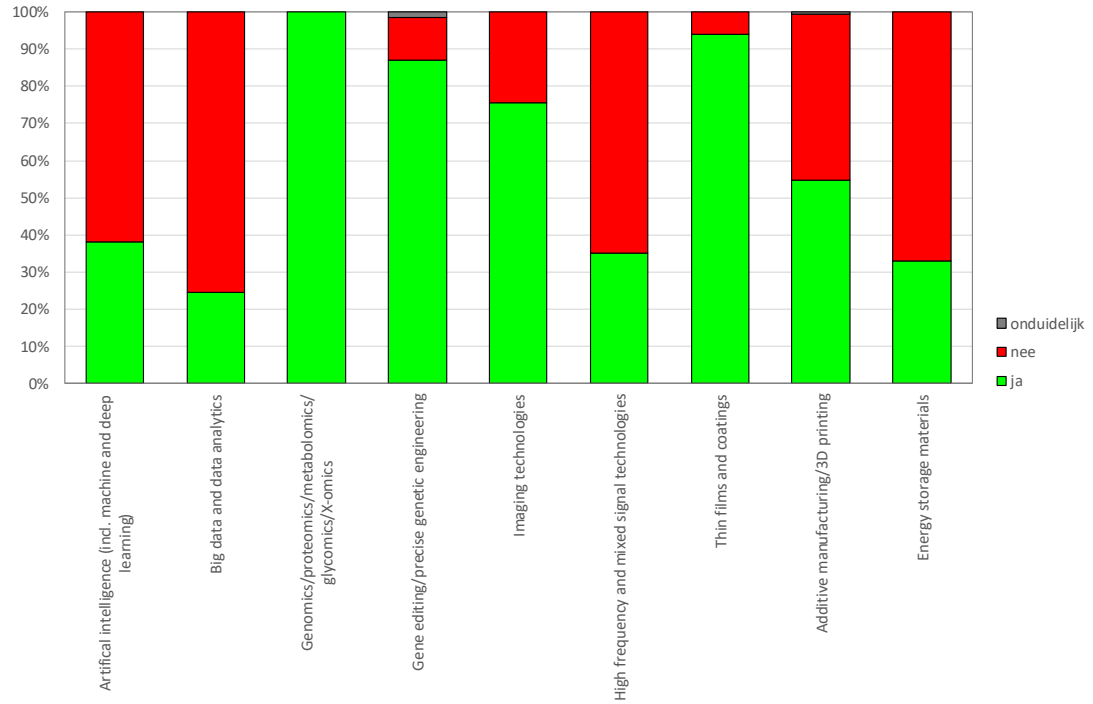
Bij de overige vier sleuteltechnologieën werd het overgrote deel van de projecten beoordeeld als correct toegewezen. In drie gevallen komt het percentage dat correct is toegewezen boven de grens van 80 procent *precision*.
- Gene editing/precise genetic engineering (87 procent "ja").
- Genomics/proteomics/metabolomics/glycomics/X-omics (100 procent "ja").
- Imaging technologies (76 procent "ja").
- Thin film and coatings (94 procent "ja").

Wanneer we een onderscheid maken tussen projecten met een sterke, gemiddelde of zwakke koppeling met een sleuteltechnologie (figuren 5, 6 en 7), dan valt het volgende op:
- Voor acht van de negen sleuteltechnologieën is het grootste deel van projecten met een sterke koppeling correct toegewezen (met "Energy storage materials" als laagste met 67 procent "ja"). In vijf gevallen komt het percentage dat correct is toegewezen boven de grens van 80 procent *precision*.[7] Alle projecten met sterke koppeling die aan "Gene editing/precise genetic engineering", "Genomics/proteomics/metabolomics/glycomics/X-omics", "Imaging technologies", "Thin film and coatings" zijn volgens de adviseurs correct toegewezen. Bij drie sleuteltechnologieën komt het percentage dat correct is toegewezen zelfs bij de projecten met een sterke koppeling niet boven de 80 procent.
- Bij "Big data and data analytics" is slechts 22 procent van de projecten met een sterke koppeling correct toegewezen.
- Er is weinig verschil in uitkomsten tussen de projecten met een gemiddelde of een zwakke koppeling met een sleuteltechnologie.
- Bij de projecten met een gemiddelde of zwakke koppeling met een sleuteltechnologie zien we hetzelfde onderscheid tussen de vier sleuteltechnologieën die vooral correct zijn toegewezen ("Gene editing/precise genetic engineering", "Genomics/proteomics/metabolomics/glycomics/X-omics", "Imaging technologies", "Thin film and coatings") en de overige vijf sleuteltechnologieën.

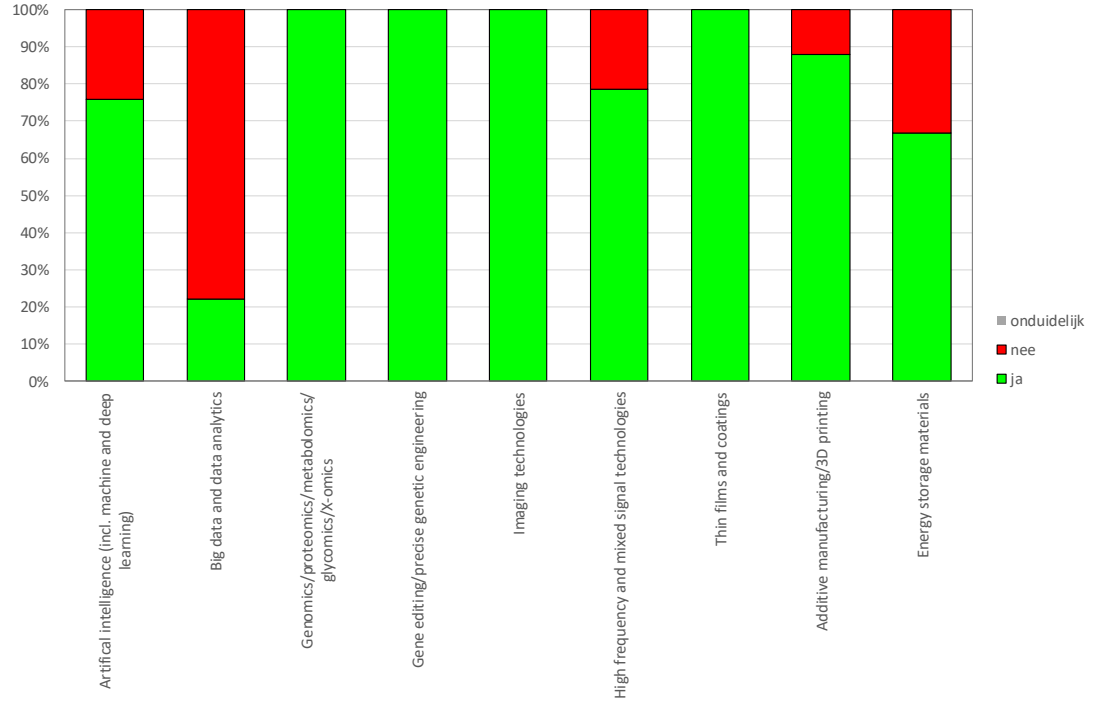
De WBSO-adviseurs merken met betrekking tot "Gene editing/precise genetic engineering", "Genomics/proteomics/metabolomics/glycomics/X-omics", "Thin film and coatings" op dat veel zoektermen ontbreken. Ten aanzien van "Energy storage materials" gaf de WBSO-adviseur aan dat veel technologieën batterijen gebruiken zodat de zoektermen teveel projecten (false positives) opleveren. Het is volgens de betreffende expert voor deze sleuteltechnologie beter om alleen in de projecttitel te zoeken.

[7] Tijdens de expertsessie werd verwezen naar de 80/20 regel: voor een goed resultaat moet tenminste 80 procent precies zijn wat wordt gezocht en mag hooguit 20 procent een 'false positive' zijn.

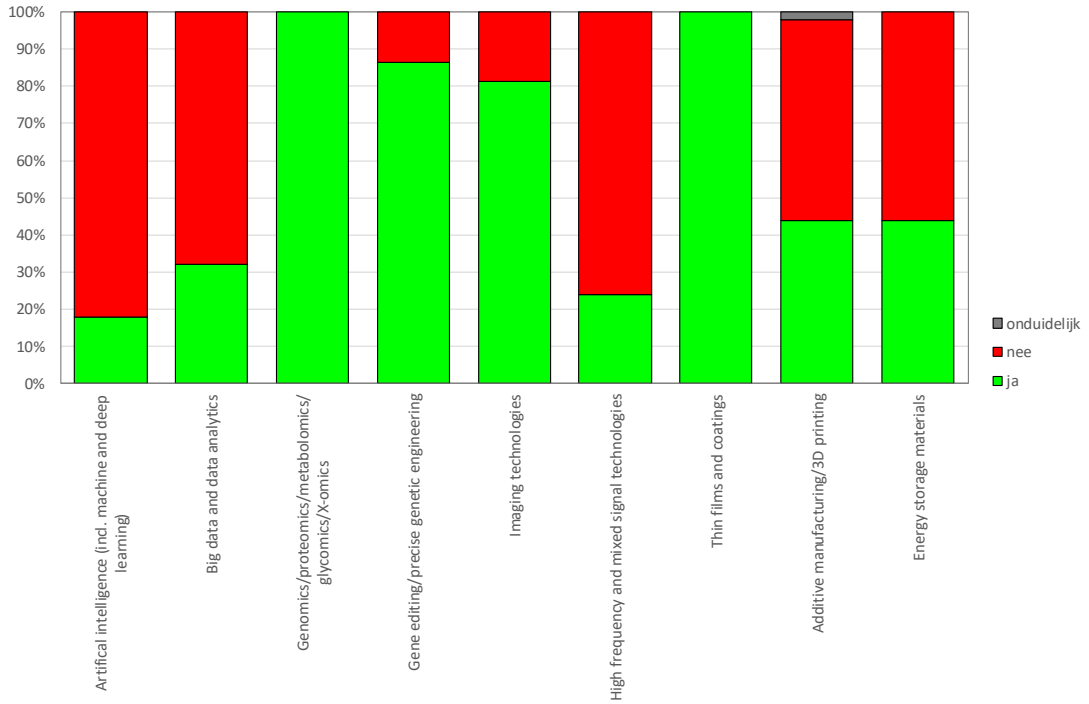
Figuur 4. Percentage van alle beoordeelde projecten dat door WBSO-adviseurs is gevalideerd (ja = correct toegewezen; nee = niet correct toegewezen)



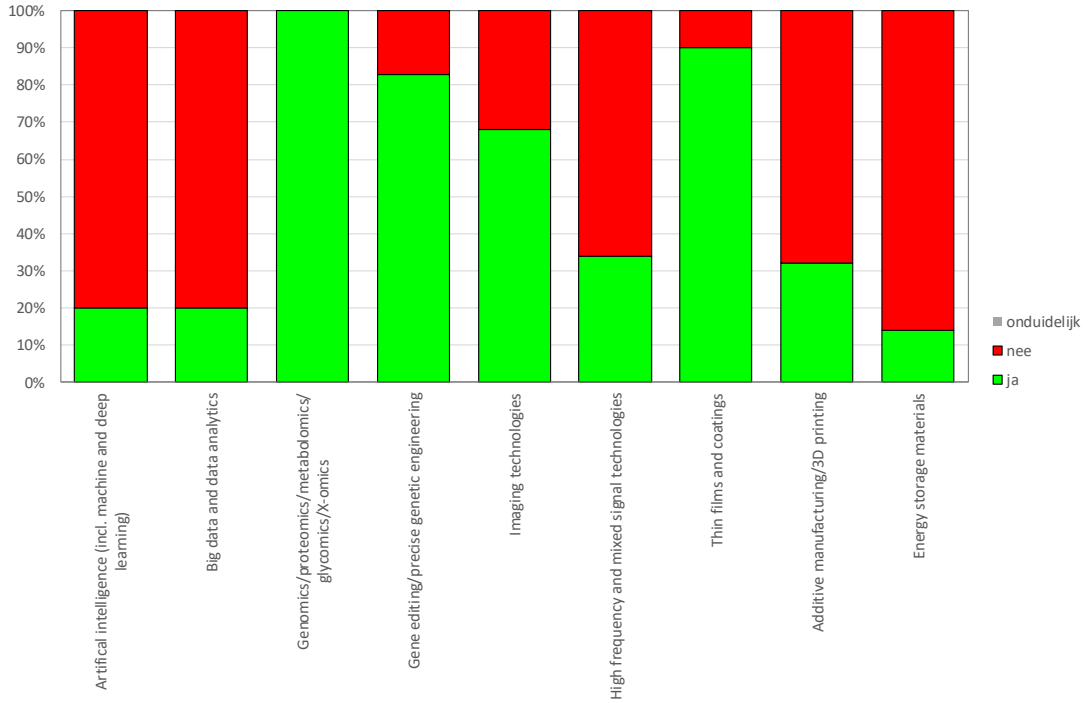
Figuur 5. Percentage van de beoordeelde projecten met een sterke koppeling dat door WBSO-adviseurs is gevalideerd (ja = correct toegewezen; nee = niet correct toegewezen)



Figuur 6. Percentage van de beoordeelde projecten met een gemiddelde koppeling dat door WBSO-adviseurs is gevalideerd (ja = correct toegewezen; nee = niet correct toegewezen)



Figuur 7. Percentage van de beoordeelde projecten met een zwakke koppeling dat door WBSO-adviseurs is gevalideerd (ja = correct toegewezen; nee = niet correct toegewezen)



De voornaamste conclusies na de validatieronde zijn:
- Er is een duidelijk verschil tussen de resultaten voor projecten met een sterke koppeling en projecten met een gemiddelde of zwakke koppeling. In het eerste geval is een groter percentage beoordeeld als correct toegewezen.
- De zoektermen lijken beter te werken voor sommige sleuteltechnologieën dan voor andere. Een relatief hoog percentage van de projecten in "Gene editing/precise genetic engineering", "Genomics/proteomics/metabolomics/glycomics/X-omics", "Imaging technologies" en "Thin film and coatings" is door de experts als "correct toegewezen" beoordeeld. Omgekeerd is een relatief hoog percentage van de projecten in "Big data and data analytics" als "verkeerd toegewezen" beoordeeld.
- Een hoog percentage "correct toegewezen" betekent niet zonder meer dat de zoektermen en de methode goed zijn. Voor "Gene-editing/precise genetic engineering" gaf de expert aan dat alle projecten correct zijn toegewezen, maar ook dat veel wordt gemist omdat veel zoektermen ontbreken. Met andere woorden, de *precision* is hoog (wat wordt gevonden is correct), maar de *recall* is laag (we vinden veel niet). Ook bij "Genomics/proteomics/metabolomics/glycomics/X-omics" en "Thin film and coatings" ontbreken volgens de experts veel zoektermen.

6.2 Vervolganalyse na validatie van zoektermen door WBSO-adviseurs

Na de validatie door WBSO-adviseurs is opnieuw in de database met WBSO-projecten gezocht naar projecten die kunnen worden toegewezen aan de negen sleuteltechnologieën. De uitkomsten van de validatie zijn daarbij als input gebruikt.

De WBSO-adviseurs hebben voor de negen sleuteltechnologieën 542 projecten aangewezen die volgens hen correct aan een sleuteltechnologie zijn toegewezen. Bovendien hebben de WBSO-adviseurs in de validatieronde 127 nieuwe zoektermen (zie bijlage 5) aangedragen.

Allereerst is een nieuwe lijst met zoektermen gemaakt. Deze lijst bestaat uit drie onderdelen:
1. de zoektermen van Elsevier (n=386);
2. de zoektermen die door de WBSO-adviseurs zijn aangedragen in de validatieronde (n=133);
3. zoektermen uit de teksten van de 542 projecten die in de validatieronde door de experts zijn gevalideerd (n=450).

De zoektermen uit de 542 gevalideerde projecten zijn uit de tekst van de projectbeschrijvingen afgeleid met behulp van een zelf geschreven tool die de Natural Language Toolkit (NLTK) van Python gebruikt (voor de source code van de betreffende modules zie bijlage 6). Met de NLTK is gezocht naar *n-grams* in de tekst van alle projectbeschrijvingen. Een *n-gram* is in deze studie gedefinieerd als een opeenvolgende reeks van woorden in een tekst. Daarbij zijn woorden die geen betekenis hebben weggelaten – de zogenaamde stopwoorden, zoals "de", "belangrijk", of "van". In totaal bevatten de 148.892 WBSO-projectbeschrijvingen ongeveer 33 miljoen *n-grams*. Dit betreft drie soorten *n-grams*:
- *unigrams*: individuele woorden, zoals "3D-printen", "data" of "enzym"
- *bigrams*: combinaties van twee woorden die voorkomen binnen een afstand van drie woorden, zoals "big data" of "internet things" (voor "Internet of Things").
- *trigrams*: combinaties van drie woorden die voorkomen binnen een afstand van vier woorden, zoals "next generation sequencing" of "hematopoietic stem cell".

Bijlage 3 laat zien hoeveel projecten zijn gevonden met de 'concepts or keywords' van Elsevier, met de zoektermen van de WBSO-adviseurs en met de ngrams uit de WBSO-projectbeschrijvingen. Daarbij wordt getoond hoeveel projecten alleen met een van deze sets zoektermen werd gevonden (bijvoorbeeld alleen met de Elsevier 'concepts or

keywords' en niet met de zoektermen van de WBSO-adviseurs of de ngrams) en hoeveel met combinaties van deze drie sets.

In de eerste fase van de studie is de sterkte van de koppeling tussen een project en een sleuteltechnologie gemeten vanuit het aantal en de frequentie van de zoektermen die werden gevonden. Toen werd gevonden dat slechts een klein deel van alle projecten (0,7 procent) een sterke koppeling heeft. Onduidelijk bleef met welke precisie de zoektermen van Elsevier een sleuteltechnologie aanduiden. Daarom is in dit stadium aan de WBSO-adviseurs gevraagd om de zoektermen inhoudelijk te beoordelen.

De nieuwe lijst met zoektermen is voorgelegd aan de WBSO-adviseurs. Zij hebben iedere zoekterm beoordeeld en een score gegeven (zie bijlage 5). De scores zijn:
- 0 = niet relevant, geen goede zoekterm;
- 1 = relevant maar te generiek (denk aan "detection");
- 2 = relevant en specifiek genoeg, maar op zichzelf onvoldoende;
- 3 = relevant en specifiek, op zichzelf voldoende om de sleuteltechnologie te vinden.

Op basis van de gevalideerde lijst met zoektermen, waarbij termen die door de experts waarde 0 zijn gegeven niet zijn meegenomen, is gezocht in de lijst met ruim 33 miljoen *n-grams*. De hypothese is dat we na deze validatie voor de 9 sleuteltechnologieën een sterkere set zoektermen hebben dan de oorspronkelijke lijst met zoektermen van Elsevier. We verwachten:
1. dat we daarmee meer projecten vinden dan eerst;
2. dat de verhouding tussen Nederlandstalige en Engelstalige projecten meer in balans is (was 15% vs. 50% over het geheel genomen; het aandeel van de 9 KET's zou bij beide sets vergelijkbaar moeten zijn, niet per se gelijk);
3. dat een groter percentage van de zoektermen gevonden wordt (van de Elsevier zoektermen werd ongeveer een derde gevonden);
4. dat we projecten vinden die we met de Elsevier zoektermen niet vonden (de *type-II* fouten);
5. dat de zoektermen die door experts zijn gevalideerd een sterkere koppeling met de projectomschrijvingen opleveren dan het geval was in de validatieronde (de oorspronkelijke zoektermen met behulp van de Elsevier software tool).

Tabel 6 vergelijkt de resultaten van de toepassing van de Elsevier tool op alle projecten in de WBSO-database (de validatieronde) met de resultaten van de toepassing van de zoektermen die voortkwamen uit de vervolganalyse na de validatie door WBSO-adviseurs. Deze tabel laat zien dat voor alle sleuteltechnologieën een groter aantal projecten is gevonden (een hogere *recall*). Daarnaast blijkt dat bij zes van de negen sleuteltechnologieën een groter deel van de gevonden projecten een sterke koppeling heeft en een kleiner deel een zwakke koppeling heeft met de sleuteltechnologie. Uitzonderingen zijn "Big data and data analytics", "Gene editing/precise genetic" en "Thin film and coatings". Voor "Big data and data analytics" werden weliswaar meer projecten gevonden, maar de resultaten van de vervolganalyse zijn minder sterk dan die van de validatie door WBSO-adviseurs. Het aantal gevonden projecten voor "Gene editing/precise genetic" is 33 keer zo groot als in de validatieronde, maar de koppeling tussen de projecten en de sleuteltechnologie is minder sterk.

Tabel 6. Het aantal gevonden projecten met de Elsevier software tool en met de nieuwe lijst van gevalideerde zoektermen

	projecten gevonden met de Elsevier software tool		projecten gevonden met de nieuwe lijst van gevalideerde zoektermen	
sleuteltechnologie	aantal	waarvan sterke koppeling	aantal	met een totaal gewicht meer dan 5
Additive manufacturing/3D printing	1.547	202 (13%)	1.720	400 (23%)
Artificial intelligence (incl. machine and deep learning)	2.530	414 (16%)	16.800	3.424 (20%)
Big data and data analytics	3.572	774 (22%)	4.030	66 (2%)
Energy storage materials	602	26 (4%)	4.633	881 (19%)
Gene editing/precise genetic engineering	131	34 (26%)	4.321	449 (10%)
Genomics/proteomics/metabolomics/ glycomics/X-omics	309	20 (6%)	5.611	1.183 (21%)
High frequency and mixed signal technologies	1.853	29 (2%)	25.119	4.904 (20%)
Imaging technologies	276	6 (2%)	17.642	1.857 (11%)
Thin film and coatings	119	10 (8%)	26.657	1.598 (6%)

Opmerking: Een project met een sterke koppeling met een sleuteltechnologie bevat meerdere zoektermen en heeft op tenminste een zoekterm meerdere matches. Het totale gewicht van de gevonden zoektermen is berekend door met de scores die door de WBSO-adviseurs aan iedere zoekterm zijn gegeven, namelijk: een gewicht van 0 voor score van 0, een gewicht van 1 voor score van 1, een gewicht van 5 voor score van 2, en een gewicht van 10 voor score van 3. Het is niet mogelijk om de resultaten van de twee rondes een-op-een met elkaar te vergelijken. Een totaalgewicht van meer dan 5 is vergelijkbaar met een sterke koppeling, omdat hiervoor meerdere zoektermen moeten worden gevonden en waarschijnlijk tenminste één zoekterm meer dan eens.

De voornaamste conclusies zijn:

- We kunnen concluderen dat de *recall* veel hoger is. Met de nieuwe, gevalideerde set zoektermen worden inderdaad meer projecten gevonden dan met de Elsevier zoektermen. Dat komt voor een deel door het grotere aantal zoektermen en het feit dat Nederlandse zoektermen zijn toegevoegd. De gevalideerde set is echter ook beter toegespitst op het zoeken naar de betreffende sleuteltechnologieën in de WBSO-database. Een hogere recall is echter niet zonder meer ook preciezer.
- De verhouding tussen het aantal Nederlandstalige en Engelstalige projecten is duidelijk meer in balans. De som van het aantal gevonden Nederlandstalige projecten in de negen sleuteltechnologieën (niet gecorrigeerd voor dubbeltellingen) was 9.139 projecten en is gegroeid naar 100.986, een verhouding van 11,1:1. De som van het aantal gevonden Engelstalige projecten is gestegen van 1.800 naar 4.737, een verhouding van 2,6:1.
- De meeste zoektermen worden in de tekst gevonden. Voor een aantal sleuteltechnologieën wordt een kleiner deel van de Elsevier zoektermen dan van de andere zoektermen gevonden. Dat duidt erop dat de aanvullende zoektermen voor de koppeling tussen WBSO-projecten en sleuteltechnologieën relevanter zijn dan de originele Elsevier zoektermen (die indertijd zijn opgesteld in relatie tot wetenschappelijke artikelen).
- In sommige gevallen lijkt de *precision* toegenomen. Wel moet worden opgemerkt dat het niet mogelijk is om in dit opzicht de resultaten van de validatie door WBSO-adviseurs en de vervolganalyse een-op-een met elkaar te vergelijken. Bij "Additive manufacting/3D printing", "Energy storage materials", "Genomics/proteomics/

metabolomics/glycomics/X-omics", "High frequency and mixed signal technologies", "Imaging technologies" en – in mindere mate – "Artificial intelligence (incl. machine and deep learning)" neemt het percentage projecten met een sterke koppeling aanzienlijk toe. Bij "Thin film and coatings" en "Gene editing/precise genetic engineering" neemt het percentage projecten met een sterke koppeling af maar stijgt het aantal gevonden projecten aanzienlijk. Bij "Big data and data analytics" blijkt het aantal gevonden projecten in de vervolganalyse slechts met 10 procent te zijn toegenomen ten opzichte van de validatie door WBSO-adviseurs, terwijl het aandeel van projecten dat een sterke koppeling kent met de zoektermen is gedaald.

- Belangrijk is te onderkennen dat het aandeel projecten met een sterke koppeling een eerste indicatie betreft van de *precision*. De aandelen zeggen echter nog niet in hoeverre de toedeling ook daadwerkelijk terecht is. Dit kan alleen worden getoetst door op basis van steekproeven de resultaten te controleren door de betreffende projectvoorstellen door te lezen. Deze exercitie heeft geen onderdeel uitgemaakt van voorliggend haalbaarheidsonderzoek.

7 Conclusies

Het doel van deze haalbaarheidsstudie is om te verkennen of het mogelijk is om met de methode van Elsevier en met behulp van de bijbehorende software tool en onderliggende algoritmes de inzet op sleuteltechnologieën vanuit de WBSO in kaart te brengen. Omdat deze studie een haalbaarheidsonderzoek betreft, kunnen inhoudelijke uitkomsten niet op voorhand worden gezien als betrouwbare schattingen of welke andere indicatie dan ook van de beleidsinzet op het totaal van sleuteltechnologieën, hoofdcategorieën van sleuteltechnologieën of specifieke sleuteltechnologieën.

Over het gebruik van de Elsevier software tool in de context van de WBSO

Deze haalbaarheidsstudie toetst of het mogelijk is op basis van de Elsevier methodiek, software tool en onderliggende algoritmes bruikbare en betrouwbare resultaten te kunnen produceren vanuit de database met WBSO-projecten met een bescheiden onderzoeksinspanning en met deugdelijke validatie achteraf. Op basis van het hier voorliggende haalbaarheidsonderzoek en de in hoofdstuk 5 en 6 beschreven onderzoeksresultaten kunnen we concluderen dat de toepassing van deze alternatieve methode op basis van de Elsevier software tool voor de WBSO niet tot bruikbare resultaten heeft geleid.

Onze ervaring met het gebruik van de Elsevier methode en tool is evenwel *geenszins* op te vatten als een diskwalificatie van de text mining software en onderliggende algoritmes. De tool heeft zich eerder prima bewezen in de context van wetenschappelijke publicaties en op basis van de tool zijn betrouwbare resultaten gegenereerd en gepubliceerd. Toepassing van de tool op de WBSO was dan ook het proberen waard. De ervaring heeft geleerd dat dit om verschillende redenen geen bruikbare resultaten oplevert in de context van de WBSO.

Ten eerste is gebleken dat de zoektermen die als input voor de Elsevier software tool dienen niet gericht zijn op de inhoud van WBSO-projectaanvragen. Hier speelt dat zoektermen Engelstalig zijn terwijl de overgrote meerderheid van WBSO projectaanvragen in het Nederlands is geschreven. Ook speelt dat de zoektermen zijn opgesteld ten behoeve van eerder door Elsevier uitgevoerd onderzoek waarin werd gezocht in wetenschappelijke teksten in peer-reviewed tijdschriften. Bij de projectomschrijvingen van de WBSO is een veel diverser palet in woordgebruik aanwezig dan in Engelstalige wetenschappelijke publicaties. In wetenschappelijke artikelen is aansluiting bij wetenschappelijke discourse een must om een artikel geaccepteerd te krijgen. Daarom is er meer standaardisatie in begrippen per sleuteltechnologie. In het hier voorliggende haalbaarheidsonderzoek is dit (deels) ondervangen door WBSO-adviseurs voor negen sleuteltechnologieën te vragen aanvullende trefwoorden te genereren die specifiek zijn gericht op (meer toepassingsgerichte dan puur wetenschappelijk georiënteerde) R&D-projecten. Met de juiste zoektermen worden meer WBSO-projecten gevonden die mogelijk tot een sleuteltechnologie behoren. In deze benadering blijft validatie door experts echter nodig, zowel voor het opstellen van de zoektermen als voor het toewijzen van projecten aan sleuteltechnologieën.

Ten tweede zijn ook de algoritmes van de Elsevier software tool gevoelig voor taal. Het lemmatiseren van teksten (deels gebaseerd op openbare algoritmes en deels gebaseerd op eigen ontwikkelwerk van Elsevier) is enkel mogelijk voor Engelstalige teksten. Van alle WBSO-projectomschrijvingen is echter slechts 4 procent Engelstalig en 96 procent Nederlandstalig. Text mining bevindt zich momenteel nog in een pioniersfase, waardoor openbare text mining modules nog niet wijdverbreid zijn. Nederlandstalige modules die

zich op een breed vlak hebben bewezen zullen waarschijnlijk de komende jaren nog niet vrij (of tegen redelijke kosten) verkrijgbaar zijn.

Over het verdere potentieel van text mining in de context van de WBSO

Ondanks de uitkomst dat deze alternatieve text mining methode voor de WBSO geen bruikbare resultaten oplevert zijn wel een aantal waardevolle lessen geleerd. Het werk dat in het afgelopen halfjaar is gedaan heeft nuttige resultaten en inzichten opgeleverd ten aanzien van de mogelijkheden en onmogelijkheden bij eventuele vervolgstappen. Ondanks dat een toepassing van de Elsevier methode en software tool op WBSO-projectomschrijvingen niet tot bruikbare inhoudelijke resultaten heeft geleid zien we wel degelijk potentie voor de toekomst om op basis van text mining thematische analyses op de WBSO uit te kunnen voeren.

Voor de korte termijn zullen we terug moeten vallen op de eerder toegepaste methode van gewogen trefwoordenanalyse. De nadelen bij deze methode (met name de validatie kent een vrij arbeidsintensief karakter) nopen echter om te zoeken naar alternatieve text mining methoden. In een expertoverleg (bijlage 4) is de potentie van de text mining strategie van *supervised machine learning* benadrukt. Bij deze methode worden categorieën niet gedefinieerd op basis van vooraf gekozen trefwoorden, maar worden een aantal specifiek voor het domein relevante projecten per categorie geselecteerd. De text mining software haalt vervolgens uit deze projecten (die als zogenaamde trainingset dienen) de specifieke termen en combinaties van termen die maken dat het betreffende project juist tot deze specifieke categorie behoort.

Welke methode, welk model of welke software ook wordt toegepast, het toepassen van text mining voor nieuwe thema's is en blijft een arbeidsintensief proces. Iteratie en validatie (het vaststellen van de false positives en de false negatives) met behulp van experts blijft een must voor een correct resultaat. Hoewel – bij de identificatie van nieuwe thema's – het opstellen van een adequate trainingset nog steeds een arbeidsintensief proces is, is het vooruitzicht dat – als de trainingset eenmaal is opgesteld – supervised machine learning geautomatiseerd (en relatief goedkoop) kan worden opgepakt met een beperkte validatie achteraf. Voor bestaande thema's geldt dat de reeds handmatig geclassificeerde projecten (als bestaand onderdeel van reeds uitgevoerde gewogen trefwoordenanalyses) direct opgenomen kunnen worden in de trainingset. Een deel van het arbeidsintensieve voorbereidende werk is derhalve al uitgevoerd.

Bijlage 1: Sleuteltechnologieën

Groep	Sleuteltechnologie
Advanced Materials	Bio (related) materials and soft material
	Composite and ceramics
	Designer and meta materials
	Energy conversion
	Energy storage materials
	Optical/electronic/magnetic materials (incl 2D and graphene)
	Smart/self healing/self-organising materials
	Structural materials
	Thin films and coatings
Chemical technologies	(Bio)Process technology including process intensification
	Analytic technologies
	Catalysis
	Electrification / Hydrogen technology / power to gas
	Microreactors
	Separation technology
Digital technologies	Artificial intelligence (incl. machine and deep learning)
	Big data and data analytics
	Block chain
	Encryption technologies/ digital security
	High Performance Computing Grid Computing and Cloud Technologies/Computing
Engineering and fabrication technologies	(Opto)mechatronics
	Additive manufacturing/3D printing
	Cyberphysical systems
	High frequency and mixed signal technologies
	Imaging technologies
	Robotics
	Sensors and actuators
Life sciences technologies	Biocatalysis
	Biochips and biosensors
	Biofabrication
	Gene editing/precise genetic engineering
	Genomics/proteomics/metabolomics/ glycomics/X-omics
	Industrial biotechnology (white)
	Nanomedicine
	Organ on a chip
	Stem cell technology
	Synthetic cell technology
Nanotechnologies	Bionanotechnology
	Micro and nanofluidics
	Nanomanufacturing
	Nanomaterials
	Nanoscale devices
	Semiconductor devices
Photonics and light technologies	Integrated photonics
	Photon generation technologies
	Photonic detection
	Photovoltaics
Quantum technologies	Quantum communication
	Quantum computing
	Quantum sensors and metrology

Bijlage 2: Elsevier's zoektermen per sleuteltechnologie

(Bio)Process technology including process intensification
Batch Fermentation
Bioethanol
Biohydrogen
bioprocess
bioprocess technology
Bioprocessing
bioreactor
Bioreactors
cell factories
Cofactor Engineering
complex streams
enzymatic conversion
Ethanol Fermentation
Ethanol Production
Fermentation
Fermentation Wastes
Lactic Fermentation
Liquid State Fermentation
Metabolic Engineering
microbial conversion
microbial kinetics
multi-phase flow
Nanofiltration
Nanofiltration Membranes
particle-size distribution
Photobiological Hydrogen Production
Photobioreactors
Photofermentation
pretreatment and hydrolysis
Silage Fermentation
Solid State Fermentation
spinning disc reactor
Submerged Fermentation

(Opto)mechatronics
Adaptive Optics
Airborne Telescopes
Camera Calibration
Cameras
Computational optics
Extreme Adaptive Optics
Extremely Large Telescopes
Free-form optics
Gravitational wave telescopes
High Speed Cameras
Imaging Spectrometer
Infrared Camera
Integral Field Unit
James Webb Space Telescope
millimeter telescope
mm and submillimeter telescopes
Multi-conjugate Adaptive Optics
Multispectral Band Cameras
Opto-mechatronics
Panoramic Cameras
Shack-hartmann Wavefront Sensor
Smart Optics
Spaceborne Telescopes
Spectrometers
Spica
submillimeter telescope
Telescopes
Wavefront Sensing

Additive manufacturing/3D printing
3D model
3D model retrieval
3d Modeling
3D modelling
3d Printers
3D printing
additive manufacturing
binder jetting
Biofabrication
Bioprinting
CAM
Cam-clay Model
Computer Aided Manufacturing
Digital twinning
directed energy deposition
Electron Beam Melting
filament extrusion
fused deposition modeling
Layered Manufacturing
Manufacturing Technology
material extrusion
material jetting
Medical splints
powder bed fusion
Printing Machinery
selective laser melting
Selective Laser Sintering
Selective laser sintering
sheet lamination
Splints (medical)
Stereolithography
vat photopolymerization

Analytic technologies
analytical separation
Anaphase-promoting Complex-cyclosome apc1 Subunit
Atomic Emission Spectroscopy
chemical resolution
chemometrics
chromatography
Comprehensive Two-dimensional Chromatography
Dielectric Spectroscopy
Electric Force Microscopy
electrophoresis
Extended x Ray Absorption Fine Structure Spectroscopy
Field Emission Microscopes
Flow Cytometry
High Resolution Transmission Electron Microscopy
high-throughput
High-throughput nucleotide Sequencing
High-throughput Screening Assays
Hydrophilic Interaction Chromatography
Mass Spectrometry
mass spectroscopy
microscopy
Nuclear Magnetic Resonance
Photoacoustic Microscopy
Photoluminescence Spectroscopy
Raman Spectroscopy
Reflectance Confocal Microscopy
Scanning Electron Miscroscopy
solid state NMR
spatial resolution
spectroscopy
ssNMR
Surface Plasmon Resonance Imaging
temporal resolution
Terahertz Spectroscopy
Two-dimensional Difference Gel Electrophoresis
Ultrafast Liquid Chromatography
Ultra-performance Liquid Chromatography

Artifical intelligence (incl. machine and deep learning)
Ant Colony Optimization
artificial intelligence
Autonomous decision making
Autonomous systems
Back-propagation Neural Network
Bam Neural Networks
Boltzmann Machine
Bp Neural Network
Causal interference
Cohen-grossberg Neural Networks
Computational Creativity
Co-training
Deep Learning
Delayed Neural Networks
Echo State Network
Elman Neural Network
Extreme Learning Machine
Genetics-based Machine Learning
Machine learning
Neural networks
Neuroevolution
Probabilistic Neural Network
Pulse Coupled Neural Network
Radial Basis Function Neural Network
Recurrent Neural Networks
Reinforcement Learning
Semi-supervised Learning
Sentiment Classification
Spiking Neural Networks
Stochastic Neural Networks
Supervised Learning
Supervised Machine Learning
Swarm Intelligence
Turing Test
Unsupervised Machine Learning
Wavelet Neural Network

Big data and data analytics
Apriori Algorithm
Automatic Image Annotation
Big Data
Data accuracy
Data confidentiality
Data fairness
Data Mining
data science
data stewardship
Data transparency
Deep Learning
distributed sensor data
Efficient deep learning
Findable data
Frequent Pattern Mining
heterogeneous data
image analysis
Image Annotation
information retrieval
learning algorithms
machine learning
Mapreduce
Opinion Mining
Pattern Mining
Privacy Preserving Data Mining
Process Mining
radiomics
Responsible data

Re-usable storage of data
Sentiment Analysis
Text Analysis
Unsupervised Machine Learning
Visual Analytics
Bio (related) materials and soft material
Bioactive Glass
Bioceramics
Biocompatibility
Biocompatible Coated Materials
Biofabrication
biomaterial
Biomaterials
Biomedical Materials
biomimetic
Biomimetic Materials
Biomimetic Processes
Biomimetic Synthesis
Biomimetics
Biophysics
Bioprinting
Cell Engineering
colloids
Nanocarriers
Nanoconjugates
soft material
Tissue Engineering
Tissue Scaffolds
Biocatalysis
Asymmetric synthesis
Biocatalysis
Biotransformation
Chemoselectivity
Chiral
Diastereoselectivity
Directed Evolution
Enantiomers
Enantiopure compounds
enantioselectivity
Enzyme
Enzyme engineering
Enzymes as Catalysts
Hydrolase
Immobilized enzymes
Industrial enzymes
Isomerase
kinetic resolution
Ligase
Lyase
natural catalysts
Nitrilase
Oxalate Decarboxylase
Oxidoreductase
Protein Engineering
Protein-ligand interactions
Reaction Engineering
Regioselectivity
Superoxide Reductase
Transferase
Biochips and biosensors
biochip
Biochips
Bioluminescence Resonance Energy Transfer Techniques
biosensing
biosensor
Biosensors
biosignal
Chemical Detection
Digital Microfluidics
Electrochemical Sensors
Electronic Transistors
Enzyme Sensors
Fluidic Devices
Glass Membrane Electrodes
Glucose Sensors
Immobilized nucleic Acids
Immobilized Proteins
Immunosensors
Integrated Nanoliter Systems
lab-on-a-chip
Lab-on-a-chip Devices
Microbial Detection
Microfluidic Analytical Techniques
microfluidics
Molecular Imprinting
Nanosensors
Nucleotide Aptamers
Optofluidics
Single-cell Analysis
Biofabrication
3D biofabrication
3d Reconstruction
Bioactive Glass
Bioartificial Organs
Biocompatibility
Biocompatible Coated Materials
Biofabrication
biofrabrication
biological models
Biomaterials
Biomimetic Processes
Bioprinting
Bone Regeneration
Cell Engineering
cell on a chip
Cellular Microenvironment
Gas Foaming
Guided Tissue Regeneration
Induced Pluripotent Stem Cells
organ on a chip
Patient-specific Modeling
Polycaprolactone
Regenerative Medicine
Scaffolds
tissue constructs
Tissue Engineering
tissue on a chip
Tissue Regeneration
Tissue Scaffolds
Tricalcium Phosphate
Bionano
Biophysical Phenomena
Biophysical Processes
Biophysics
Biosensing
Biosensors
Biotransducer (electro chemical / ion switch / fluorecent)
Cellular Biophysics
cytoskeleton
Electro chemical transducer
Electronic Transistors
fluorecent transducer
ion switch transducer
magnetic tweezers
microfibrils
Molecular Biophysics
Nano-molecular machines
nanopores
optical tweezers
single molecule techniques
Block chain
Autonomous system
Autonomous Systems
BaaS
Bitcoin
Bittorrent
blockchain
blockchain as a service
Buck Converter
consensus algorithm
Content Distribution
cryptocurrency
Cryptology technologies
Digital trust
Distributed ledger
distributed ledger technology
distributed trust
Electronic Money
Ethereum
File Sharing
Free Riders
hash mining
Iota cryptocurrency
litecoin
Monero
Nonautonomous Systems
Overlay Networks
P2p
P2p Network
P2p Systems
Peer to Peer
Peer-to-peer (p2p)
Peer-to-peer Computing
Peer-to-peer Networks
Peer-to-peer Systems
permissionless
Reputation Management
Reputation System
Ripple
Ripple Effect
Self-driving economy
smart contracts
Switching Frequency
Trust Model
unpermissioned
Catalysis
adsorption
Alkenylation
Allophanate Hydrolase
Allylation
Benzene Refining
Biocatalysis
Catalysis
catalyst
Catalyst activity
catalyst regeneration
Catalyst selectivity
Catalyst supports
catalytic
Catalytic Oxidation
Chemical activation
Cross Coupling
dehydrogenation
Diastereoselectivity
Electrocatalysis
Enantioselectivity
enzyme
Enzymes
Fischer-Tropsch
Fluoroamines
Green Chemistry Technology
Hydrogenation
ligand
Ligands
Michael Reaction
Nanoreactors
Olefins
Palladium
Photocatalysis
Photocatalysts
Photochemical Processes
photochemistry
polymerization
Reaction Intermediates
Regioselectivity
zeolite
Composite and ceramics
Barium Zirconate
Bioceramics
ceramic composites
Ceramic Compositions
Ceramic Foams
Ceramic Honeycombs
ceramic matrix composites
Ceramic Petrography
Ceramic Production
Ceramic Technology
Ceramic Traditions
Ceramics

ceramics alloys
ceramics and composites
Chinese Ceramics
Composite Ceramic
Composite ceramics
composite material
Composite materials
composite structure
composites and ceramics
Ferroelectric Ceramics
Glass Ceramics
Inca Period
Lead Zirconate Titanates
metal matrix composites
metal matrix composites
Metal Oxide Ceramics
Niobium Oxide
Organic Residue Analysis
Organically Modified Ceramics
Oxyfluorides
Piezoceramics
Piezoelectric Ceramics
polymer composites
Semiconducting Lead Compounds
Sintered Alumina
Structural Ceramics
thermoplastic composites

Cyberphysical systems

Abs Algorithms
Arm Processors
automatic pilot avionics
autonomous automobile systems
autonomous system
Autonomous Systems
Communication Networks
computer architecture
control algorithm
cybernetics
cyberphysical systems
cybersecurity
Data leakage
data security
Database intrusion detection
Database monitoring
DDos
Denial of Services
Denial-of-service Attack
Design Space Exploration
Distributed Control
distributed system security
Dynamical Systems
Economic Cybernetics
Embedded Processor
Embedded Software
embedded system
Embedded Systems
encryption
Generalized Predictive Control
Gpgpu
Hardware Accelerator
Hardware/software Co-design
Information Exchange
information security
Insider threats
Instruction Sets (computers)
internet of things
Iterative Learning Control
known-plaintext attack
M2M communication
machine-to-machine communication
Many-core
Memory Architecture
Memory Hierarchy
Memory Management Units
Microarchitecture
Model driven design
monitoring algorithm
Mpsoc
Multi-core
Multi-core Processor
Multicore Programming
Network Architecture
Network on chip
Nonautonomous Systems
Open Architecture
Organizational Cybernetics
process control systems
Real-time and Embedded Systems
Reconfigurable Architectures
robotics systems
Second-order Cybernetics
signature verification
smart grid
system-on-a-chip
Transactional Memory
trust management
Viable System Model
Wireless Interconnects
wireless sensor

Designer and meta materials

Amphiphiles
anisotropic transformation optics
Assembling
atomic scale manipulation
Biofabrication
Biomimetic Materials
Bioprinting
Block Copolymers
chiral Metasurfaces
chiro-optical metasurface
cloaking device
Contact Hole
Defectivity
Designer and meta materials
functional complex matter
Graphoepitaxy
Hydrophobins
integration within device architecture
Invisibility Cloaks
magneto-optic effect
materials synthesis
metamaterial
Metamaterial cloaking
Metamaterials
Microphase Separation
negative poisson ratio
Negative Refraction
Negative refraction index
negative refractive index
Peptide Nanotubes
Radar Absorbers
Ring Gages
Ring Resonator
Self Assembly
self-assembly material
Stealth Technology
Zero Property

Electrification / Hydrogen technology / power to gas

Alkaline Fuel Cells
Auxiliary Power Sources
Catholytes
Cell Anodes
Cell Cathodes
Direct Alcohol Fuel Cells (dafc)
Direct Borohydride Fuel Cells (dbfc)
Direct Carbon Fuel Cells (dcfc)
Direct Ethanol Fuel Cells (defc)
Direct Methanol Fuel Cells (dmfc)
Electrocatalysts
electrochemical conversion of carbon dioxide
electrochemical reduction
electrochemical synthesis
electrochemistry
Electrolytic Reduction
energy carriers
energy storage in chemical bonds
Formic Acid Fuel Cells (fafc)
fuel cell
Fuel Cell Power Plants
Fuel Cells
Fuel Storage
Gas Fuel Purification
h-2 production
h-2 storage
heterogeneous catalysis
Hydrogen Fuel Cells
hydrogen production
hydrogen storage
hydrogen technology
Hydrogen-based Energy
Methanol Fuels
Molten Carbonate Fuel Cells (mcfc)
PEM
Phosphoric Acid Fuel Cells (pafc)
photocatalysis
power to gas
Proton Conductivity
proton exchange membrane
Proton Exchange Membrane Fuel Cells (pemfc)
Protonic Ceramic Fuel Cells (pcfc)
redox catalysis
redox flow battery
redox-flow
Regenerative Fuel Cells
renewable energy
smart grid
SOFC
solar fuel
solid oxide fuel cell
Solid Oxide Fuel Cells (sofc)

Encryption technologies/ digital security

attribute based credentials
Common Divisor
Diffie-hellman
Discrete Logarithm Problem
elliptic curve cryptography
Elliptic Curve Cryptosystem
Elliptic Curves
encryption technology
Fault Attacks
Fault Injection
Finite Field Arithmetic
Gentry
Homomorphic Encryption
Montgomery multiplication
Multiparty Computation
multi-party computation
Mutual Authentication
Non-malleability
Oblivious Transfer
Pairing-based Cryptography
Post quantum crypto
Privacy Preservation
Privacy Preserving
Privacy Preserving Data Mining
Private Information Retrieval
Public Key Cryptography
Scalar multiplication
Secure Computation
Secure multi-party Computation
Side Channel Attack
Side Channel Attacks
Side-channel Analysis
side-channel cryptanalysis
Universal Composability

Zero-knowledge
Zero-knowledge Proof
zero-knowledge proofs
Energy conversion
Aboveground Biomass
Affordable gas
alkaline electrolyse
biomass
Biomass Burning
Biomass Power
Brown Carbon
Compact conversion and storage techniques
Compressed Air Energy Storage
Cryogenic Energy Storage
Electric Energy Storage
Energy Conversion
Energy Conversion Efficiency
Energy Storage
Flywheel Propulsion
Heat Storage
Hydrochars
Molten salt
Molten Salt nuclear Reactors
Power to Chemicals
Power to Gas
Power to Heat
Power to Hydrogen
solar conversion
solar energy conversion
solar to electrical
solar to electricity
solar to fuel
Solar to Gas
solar to hydrogen
Torrefaction
Wave Energy Conversion
Energy storage materials
Alkaline Batteries
Alkaline Battery
batteries
Battery
Battery Chargers
Battery Disposal
Battery Electric Vehicles
Battery Management Systems
Button-cell Battery
CAES
CAES diabatic
Catholytes
Charging (batteries)
Compressed Air Energy Storage
Compressed Air Energy Storage (CAES)
DC compressor
direct current compressor
Electric Battery
Electric Current Collectors
Electrochemical Cells
electrochemical condensator
electrochemical storage
electromechanical storage
Energy Storage
energy storage material
Flow Batteries
flywheel energy storage
Flywheel Propulsion
Fuel Cells
Inductive Power Transmission
Ion Storage
Iron Phosphates
kinetic energy storage
Lead Acid Batteries
Lithium Batteries
Lithium Compounds
Lithium Deposits
Lithium Sulfur Batteries
Lithium-ion Batteries
Metal Air Batteries
Miniature Batteries
Nickel Metal Hydride Batteries
Nuclear Batteries
Plug-in Electric Vehicles
Power-to-power (P2P)
Primary Batteries
Pulse Charging
pumped hydroelectric storage (PHS)
Shell Anodes
Silicon Batteries
Superconducting Magnetic Energy Storage (SMES)
Thermal Batteries
Thin Film Lithium Ion Batteries
Zinc-bromide Batteries
Zinc-oxygen Batteries
Gene editing/precise genetic engineering
Cas9
Clustered Regularly Interspaced Short Palindromic Repeats
Cofactor Engineering
Comparative Genomics
CRISPR
Crispr-associated Proteins
Crispr-cas Systems
Dna End-joining Repair
gene editing
gene therapy
gene transfer
gene transfer techniques
Genetic engineering
Genetic Modification
genome analysis
Genome Assembly
Genomic Structural Variation
Guide Rna
Inverted Repeat Sequences
Lentivirus
Metabolic Engineering
Molecular Farming
Molecular Sequence Annotation
Nuclear Transformations
Targeted Gene Repair
Genomics/proteomics/metabolomics/ glycomics/X-omics
bioinformatics
Comparative Genomics
epigenomics
Exome
Functional Genomics
genome analysis
Genome Assembly
genomic engineering
Genomic Structural Variation
genomics
Glycomics
High-throughput nucleotide Sequencing
Metabolome
metabolome analysis
Metabolomics
Metagenomics
Molecular Sequence Annotation
National Human Genome Research Institute (u.s.)
next generation sequencing
proteome analysis
Proteomics
RNA sequencing
Transcriptomics
Two-dimensional Difference Gel Electrophoresis
Unigenes
whole exome sequencing
High frequency and mixed signal technologies
5G
5G mobile communication systems
Antenna feeders
Antenna phased arrays
aperture arrays
Beamforming
Bolometers
CMOS
Coplanar Waveguides
Cryogenics
Demodulation
Detection
Electromagnetic Wave Polarization
Focal Plane Arrays
Frequency Bands
Frequency Modulation
Frequency Response
Frequency Selective Surfaces
HEB technology
Heterodyne
Heterodyne technology
High Frequency
high frequency signal technology
hot electron bolometer
Infrared Detectors
Lens Antennas
low noise amplifiers
Low power converter
Massive MIMO
Metamaterials
Microbolometer
Microwave Devices
Microwave KID technology
Microwave Resonators
Mixed signal integrated circuit
mixed signal technology
MOSFET
Multiple-input multiple-output (mimo) Systems
Noise Temperature
Phase Modulation
phased array
phased array feed
Q Factor
Radio Astronomy
Responsivity
Sidebands
Submillimeter
Submillimeter Waves
superconducting detectors
Superconducting Devices
Superconducting Resonators
Surface Plasmon
Terahertz
Terahertz Imaging
Terahertz Radiation
Terahertz Spectroscopy
Terahertz Wave Detectors
Terahertz Waves
TES technology
Thermal Energy
THz sensing
Titanium Nitride
transition edge sensors
Tunable
Tunnel Junctions
Ultra-wideband (uwb)
Waveguides
High Perfomance Computing Grid Computing and Cloud Technologies/Computing
Big Data
Bittorrent
Cloud Computing
Cloud Model
cloud technology
Dark Silicon
Data as a Service (daas)
Data Centre
Distributed Algorithms
Distributed computing
Distributed Data

Distributed Data Mining
Distributed Database Systems
File Sharing
Gpgpu
Hadoop
high performance computing cluster
HPCC
Mapreduce
Mobile Cloud Computing
P2p
P2p Systems
Parallel Architectures
Parallel computing
Peer-to-peer
Peer-to-peer Networks
Service Level Agreement
Software-as-a-service
Storage as a Service (staas)
Virtual Machine
Virtualization

Imaging technologies

Bolometers
Cardiac-gated Imaging Techniques
Cardiac-gated Single-photon Emission Computer-assisted Tomography
Computer Assisted Image Interpretation
Computer Assisted Radiographic Image Interpretation
Cone-beam Computed Tomography
Diffuse Optical Tomography
Diffusion Tensor Imaging
Discrete Tomography
Dual Energy Scanned Projection Radiography
Electric Impedance Tomography
Electrical Capacitance Tomography
Electrical Impedance Tomography
Electrical Resistivity Tomography
Electron Microscope Tomography
far-infared imaging
Fiducial Markers
Filtered Backprojection
Four-dimensional Computed Tomography
Free-form Deformation
Geometric Tomography
Image Encryption
Image Inpainting
image processing
image reconstruction
Image Registration
image-guided intervention
Image-guided Radiotherapy
Imaging technology
Infrared Detectors
infrared imaging
Lens Antennas
Medical Image Processing
Medical imaging
Micro-ct
Microresonators
Microwave Resonators
molecular imaging
Multidetector Computed Tomography
Multimodal Imaging
Noise Temperature
Optical Coupling
Optical imaging
Optical Tomography
Phase Contrast
Photoacoustic Tomography
Positron Emission Tomography
Radio Astronomy
Radio imaging
Remote Sensing Image
Respiratory-gated Imaging Techniques
Retinal Photoreceptor Cell Inner Segment
Seismic Refraction Tomography
Seismic Tomography
Stereoscopy
Submillimeter Waves
Super-resolution
Terahertz imaging
Terahertz Radiation
Terahertz Waves
tomography
Ultrasound Image
Whole Body Imaging
X-ray imaging
X-ray Microtomography

Industrial biotechnology (white)

bio process technology
Biofortification
Clustered Regularly Interspaced Short Palindromic Repeats
Cofactor Engineering
Crispr-associated Proteins
Crispr-cas Systems
Genetic Engineering
Guide Rna
industrial biotechnology
Metabolic Engineering
Metabolic Flux Analysis
Molecular Farming
multiphase flow
nanofiltration membrane
particle-size distribution
Production microorganisms
spinning disc reactor
Synthetic Biology
white biotechnology

Integrated photonics

Analog-optical interconnection technology
Bandpass Filters
Beamforming
Brillouin Scattering
Electric Delay Lines
Fiber Optics Communications
Integrated Optics
Integrated photonic smart antennas
integrated photonics
Mach-zehnder Interferometers
microphotonics
Microwave Filters
Microwave Frequencies
Microwave Photonics
Multimode Interference
Nanophotonics
Notch Filters
Optical Fiber Communication
Optical Resonators
optical signal technology
Phase Modulation
Phase Shifters
photonic applications
photonic beamforming
photonic chips
Photonic Devices
Photonic Integrated Circuits
Photonic Integration Technology
photonic phased array system
photonic signal processing
Photonics
Polarization Modulation
Reconfigurability
Ring Resonator
Sidebands
Silicon Photonics
Stimulated Brillouin Scattering
Tunable Filter
Waveguide Filters

Micro and nanofluidics

Drop Transfer
Fluidic Devices
Lab-on-a-chip
Lab-on-a-chip Devices
microchannel
Microchannel Plate
Microchannels
Microchip Electrophoresis
Microfluidic Analytical Techniques
Microfluidics
microfluids
Microreactor
nanochannel
Nanofluidics
Nanofluids
Optofluidics

Microreactors

bench-top microreactor
Brinkman number
Cell Surface Display Techniques
Continuous flow reactors
Dried Blood Spot Testing
Drug Discovery
Electronic Cooling
flow chemistry
Fluidic Devices
Heat Sinks
High Throughput
high throughput screening
High-throughput Screening
High-throughput Screening Assays
Image Storage Tubes
in-line analysis
lab on a chip
Lab-on-a-chip
Lab-on-a-chip Devices
Laboratory Automation
micro process engineering
Microchannel
Microchannel Plate
Microchannels
Microfluidic Analytical Techniques
Microfluidics
Microreactor
millireactors
novel process windows
process intensification
process on a chip
reaction telescoping
Screening Experiment
Slip Flow
Slug Flow
T reactor

Nanomanufacturing

Computational Lithography
Defectivity
Double Patterning
E-beam Lithography
Electroepitaxy
Electron Beam Lithography
epitaxy
Euv Mask
Euv Source
Extreme Ultraviolet Lithography
Immersion Lithography
Ion Beam Lithography
Lithography
Lithography Simulation
Mask Inspection
Maskless Lithography

Metallorganic Vapor Phase Epitaxy
Molecular Beam Epitaxy
Molecular Resist
Nanoimprint
Nanoimprint Lithography
Nanolithography
Nanomanufacturing
Photomasks
Source Mask Optimization
thin film deposition
Vapor Phase Epitaxy

Nanomaterials

alumina
Atomic layer chemical vapor deposition
atomic layer deposition
C60
Carbon nanotubes
Dendrimer
divanadium pentaoxide
fullerene
Gold nanoparticle
Gold Nanoparticles
Graphene
Graphene Devices
Graphene Transistors
Iron nanoparticles
Magnetic Nanoparticles
medical nanotechnologies
Medical Nanotechnology
Metal Nanoparticles
Multi-walled carbon nanotube
Nanocarriers
nanocatalyst
nanoceramic
Nanoclay
Nanoclays
nanocolloidal material
nanocomposite
nanocrystal
nanofiber
Nanoflowers
Nanofluids
nanofoam
nanohazard
nanolayer
Nanomagnetics
nanomaterial
Nanomedicine
nanometal
nanoparticulate material
nanophotonic material
nanopolymer
nanopore
nanoribbon
nanorod
nanosafety
Nanosheets
nanostructured coating
nanostructured film
nanostructured surface
nanotextured surface
nanotube
quantumdot
Silica Fume
Silver Nanoparticles
Single-walled carbon nanotube
Strontium titanate
strontium titanium trioxide
Theranostic Nanomedicine
Tio2
Titanium Dioxide
titanium dioxide nanoparticle
zeolite

Nanomedicine

Adeno-associated Virus 6
Albumin-bound Paclitaxel
biomarkers
gene delivery
Gold Nanoparticles
Gpi-anchored Folate Receptors
in vivo imaging
Lab-on-a-chip
Lab-on-a-chip Devices
Localized Surface Plasmon Resonance
Magnetic Nanoparticles
Magnetite Nanoparticles
Medical Nanotechnology
Metal Nanoparticles
molecular diagnostics
Nano drug delivery
Nanocapsules
Nanocarriers
Nanoconjugates
Nanofluids
Nanomagnetics
Nanomedicine
Nanoparticles
Nanoprobes
Nanorobots
nanosensors
Pharmacological Biomarkers
Protein Corona
Rnai Therapeutics
Silver Nanoparticles
Theranostic Nanomedicine

Nanoscale devices

mesodevice
mesoscale device
mesostructures
nanodevice
Nanoflowers
Nanomaterials
Nanoneedles
nanoscale device
Nanostructure Growth
nanostructures
Nanostructures (devices)

Optical/electronic/magnetic materials (incl 2D and graphene)

2D material
2D materials
antiferromagnet
Boron Nitride
condensed matter
Crystal Filters
dirac fermion
Electrochemical Capacitors
Electrolytic Capacitors
electronic materials
Electronic Transistors
ferromagnetism
Glassy Carbon Electrode
graphane
graphene
Graphene Devices
Graphene Oxide
Graphene sensors
Graphene Transistors
Graphite Electrodes
heterostructure
Holey Fibers
Landau level
magnetic dynamics
magnetic films
magnetic materials
magnetic thin film
magnetization dynamics
magneto-optic
magnon
Metal Air Batteries
Molybdenum Disulfide
MoS2
nanophotonics
Nanosheets
N-p-n Junctions
optical materials
optomagnetism
phonon
Photonic Band Gap
Photonic Crystal Fibers
Photonic Crystals
photonics
plasmonic device
plasmonic nanoparticle
plasmonic nanowire
plasmonic properties
Plasmonic resonance
plasmonic wave
P-n-p Junctions
quantum hall
Saturable Absorbers
silicene
Silicon Batteries
Slow Light
spin waves
spintronic
superconducting
superconductivity
superconductor
Supercontinuum Generation
surface plasmon-polariton
TMDC
topological insulator
Yig

Organ on a chip

Batch Cell Culture
Batch Cell Culture Techniques
Biofabrication
Bioprinting
cell culture
Cell Engineering
Cellular Microenvironment
Cellular Reprogramming Techniques
Cellular Spheroids
Digital Microfluidics
Fluidic Devices
Induced Pluripotent Stem Cells
Integrated Nanoliter Systems
Lab-on-a-chip
Lab-on-a-chip Devices
Microelectronics
microfabrication
Microfluidic Analytical Techniques
Microfluidics
microphysiological systems
organoid
organ-on-a-chip
Primary Cell Culture
Regenerative Medicine
Spheroids
Tissue Scaffolds

Photon generation technologies

active-matrix organic light-emitting diode
AMOLED
Atom Lasers
blue-light generation
cascaded emission
Chirped pulses
Coherent photon generation
Dissipative Solitons
entangled photon generation
entangled-photon generation
harmonic generation
laser
Laser Method
Laser Mirrors
Laser Power Transmission
laser wave guide
Laser-induced Breakdown Spectroscopy
LED
light emitting device
light emitting diode
light generation
microchip laser

mode locked laser
mode locking
Mode-locked Fiber Lasers
multi-photon generation
Nanophotonics
OLED
optical fiber dispersion
Optical fiber lasers
optical fibers
Optical spectral shaping
Organic Lasers
Organic light emitting diode
Passive Mode Locking
photon generation
photon generator
photon pair generation
photonic crystal
photonic microwave
Photonic microwave generation
Photonic microwave waveforms generation
photonic pair generation
plasmonics
quantum dot
quantum dots
Saturable Absorbers
Semiconductor Saturable Absorber Mirrors
single emitters
single photon emitters
single photon generation
single photon source
single-photon emission
Solar-pumped Lasers
Strontium 88
triple-photon generation
Ultrashort Pulsed Lasers
Visible Light Communication
wave form generation
waveform generation

Photonic detection

Anticoincidence Detectors
Autofluorescence
Basis Pursuit
CCD
Compressed Sensing
Compressive Sensing
Distributed Sensing
far-infared imaging
Image Sampling
infrared detectors
Integral Field Unit
Integral field units
Inverse Synthetic Aperture Radar
laser radiometry
Low-rank Matrices
Matching Pursuit
Matrix Completion
Message-passing Algorithms
Nirspec
optical coherence tomography
Optical Image Storage
Optical Imaging
Optical-Infrared
Photoacoustic Techniques
Photoacoustic Tomography
Photodetectors
photonic crystal cavities
photonic detection
Radar Signal Processing
Radio imaging
Random Projection
Reconstruction Algorithm
Signal Reconstruction
Signal Sampling
Sparse Approximation
Sparsity
Terahertz Imaging
Tight Frame
Total Variation Regularization
Voltage-sensitive Dye Imaging
X-ray Imaging

Photovoltaics

Absorbers (materials)
Acceptor Materials
Amorphous Silicon
Antireflection Coatings
Atomic Layer Epitaxy
Bulk Heterojunction
Cadmium Sulfide Solar Cells
Carrier Lifetime
Charge Density
CIGS
Collector Efficiency
Conductive Films
Conversion Efficiency
Copper Indium Selenides
Dielectric Materials
Dye-sensitized Solar Cells
Electron Recombination
Electron Scattering
Energy Conversion Efficiency
Film Growth
Film Preparation
Fixed Charge
Fresnel Reflectors
Gallium Oxides
Heterojunction Devices
Heterojunctions
Indium Sulfide
Maximum Power Point Trackers
Metamaterials
Multi-junction Solar Cells
Nanocrystalline Silicon
Nanoimprint Lithography
Nanophotonics
Nanowires
Open Circuit Voltage
Organic Photovoltaics
Organic Solar Cells
Oxide Films
Passivation
Perovskite Solar Cells
perovskites
Photocurrents
Photoelectrochemical Cells
Photoelectrochemical Devices
Photoelectrons
Photovoltaic
Photovoltaic Conversion
Photovoltaic System
photovoltaics
Plasmon
Plasmonics
Plasmons
Polarimetry
Polymer Solar Cells
Reactive Ion Etching
Semiconducting Selenium Compounds
Semiconductor Doped Polymers
Semiconductor Doping
Sheet Resistance
Silicon Nitride
Silicon Oxides
Solar Absorbers
Solar Azimuth
Solar Cell
Solar Cell Arrays
Solar Collector
Solar Collectors
Solar Cooling
Solar Electric Propulsion
Solar Energy
Solar Energy Absorbers
Solar Energy Technology
Solar Equipment
Solar Flux Density
Solar Power Generation
Solar Power Plants
Solar Power Station
Solar Powered Aircraft
Solar Reflectors
Solar Simulators
Solar Spectrometers
Solar-pumped Lasers
Spectroscopic Ellipsometry
Surface Plasmon Polariton
Surface Scattering
Thermophotovoltaic Conversion
Thin Film Circuits
Thin Film Flow
Thin Film Solar Cells
Thin Film Transistors
thin films
Ultrathin Films
Water Splitting

Quantum communication

Ghz State
optically coupled networks
Quantum Channel
Quantum Communication
Quantum Computation
quantum entanglement
Quantum Fisher Information
Quantum Information
quantum optics
quantum repeaters
Quantum Teleportation
Teleportation
waveguides

Quantum computing

Bogoliubov Theory
Bose-einstein Condensate
Bose-einstein Condensation
Bose-einstein Distribution
Cavity Qed
cavity quantum electrodynamics
Cloud Computing
Cluster State
Fractionalization
Gross-pitaevskii Equation
Interactive Proof Systems
Ion Traps (instrumentation)
Majorana fermions
Optical Lattices
Quantum Algorithms
Quantum Circuits
Quantum Communication
Quantum Computation
quantum computing
Quantum Electrodynamics
Quantum Entanglement
Quantum Error Correction
Quantum Information
Quantum Information Processing
Quantum Optics
Qubit
Reversible Logic
semiconductor qubits
superconducting qubits
Superconducting Resonators
topological isolators
Trapped Ions

Quantum sensors and metrology

Nanomechanical quantum systems
quantum metrology
quantum sensor

Robotics

Ambient Intelligence
Anthropomorphic Robots
artificial intelligence
automisation
autonomous systems
Chinese Room Argument
Cognitive Robotics
Collaborative robots
drone systems

Educational Robots
Evolutionary Robotics
Exoskeleton
hospital robots
Humanoid Robot
human-robot interaction
Industrial Manipulators
Input/output Logic
Interactive Narrative
Mechatronics
medical robotics
Mobile Robotics
Multipurpose Robots
Multi-robot Systems
pick and place
Qualitative Spatial Reasoning
rehabilitation robotics
Robot
Robot Programming
Robotic Cell
Robotic Surgery
Robotic Surgical Procedures
robotics
self configuration
Semi-supervised Learning
Service Robot
Simultaneous Localization and Mapping
soft robotic matter
surgical robots
Swarm Intelligence
Swarm Robotics
Turing Test
Unstructured environment
Unsupervised Machine Learning

Semiconductor devices

Aluminum Gallium Arsenide
Aluminum Gallium Nitride
Cadmium Sulfide Solar Cells
Cadmium Telluride
defect free reticles
Electronic Transistors
EUV lithography
Euv Mask
Extreme Ultraviolet Lithography
functional layers
GaN/InGaN
Iii-v Semiconductors
Ii-vi Semiconductors
Indium Antimonides
Indium Arsenide
Ingan
Layered Semiconductors
Magnetic Semiconductors
Metal Insulator Boundaries
Metal Oxide Semiconductors
microchip
Microchip Analytical Procedures
Microchip Electrophoresis
Mis Devices
Mos Devices
Narrow Band Gap Semiconductors
N-type Semiconductors
Organic Field Effect Transistors
Organic Semiconductors
Oxide Semiconductors
Passive Mode Locking
pattern fidelity
pattern immersion lithography
pellicle
Photoconductive Switches
Photonic Integration Technology
Power Semiconductor Devices
P-type Semiconductors
Quantum Dot Lasers
Quantum Dots
Semiconducting Indium
Semiconducting Tellurium
semiconductor
Semiconductor Detectors
Semiconductor Device Manufacture
Semiconductor Devices
Semiconductor Diodes
Semiconductor Doped Polymers
Semiconductor Doping
Semiconductor Industry
Semiconductor Manufacturing
Semiconductor Optical Amplifiers
Semiconductor Plasmas
Semiconductor Saturable Absorber Mirrors
Semiconductors
Source Mask Optimization
Wide Band Gap Semiconductors

Sensors and actuators

Acoustic Transducers
Deformable mirror actuators
Electronic Transducers
Flexible Substrate
High-intensity Focused Ultrasound Ablation
Integrated System
Interdigital Transducers
Magnetic Transducers
microfluidic systems
Mixed Conductive-sensorineural Hearing Loss
Photoacoustic Imaging
Photoacoustic Microscopy
Photoacoustic Techniques
Photoacoustic Tomography
Piezoelectric Transducers
Piezoresistive Transducers
Sound Transducers
transducer
Transducers
Ultrasonic Scattering
Ultrasonic Transducers
Vacuum Transducers

Separation technology

Air Purification
crystallization filter
distillation filter
extraction filter
filter membranes
flotation filter
Fuel Purification
gas filter
Gas Fuel Purification
heterogeneous mixture filter
homogeneous solution filter
liquid filter
membrane filtration
nanofiltration
Nanofiltration Membranes
separation technology
vapor filter
Water Purification

Smart/self healing/self-organizing materials

Amphiphiles
Artificial Receptors
Azobenzene
Calixarenes
Catenanes
Chain Ladder
Coordination Polymers
Crystal Engineering
Cucurbitaceae
Ionic Polymer-metal Composite
Macrocyclic Compounds
Molecular Recognition
Oxamide
Peptide Nanotubes
responsive material
reversible bonding
Rotaxanes
Self Assembly
self-assembly material
Self-healing
self-healing material
Self-healing Materials
self-organising material
self-organizing material
self-repair material
smart material
Smart Materials
stimuli responsive material
supramolecular
Supramolecular Chemistry
Viologens

Stem cell technology

Biological Engineering
Cell- and Tissue-based Therapy
cell therapy
Cellular Reprogramming
Cellular Reprogramming Techniques
Genes, Transgenic, Suicide
hematopoietic stem cell
Induced Pluripotent Stem Cells
matrigel
mesenchymal stem cell
Mitochondrial Replacement Therapy
organoid
Pluripotent Stem Cells
regenerative medicine
stem cell biology
stem cell technology
Veterinary Medicine

Structural materials

Bainite
Bainitic Steel
Bainitic Transformations
binder design
Carbon Silicon Carbide Composites
Dynamic Recrystallization
Friction Stir Welding
Grain Refinement
High Strength Alloys
Hvof Thermal Spraying
Laser Cladding
Lead Powder Metallurgy
Market Microstructure
Martensitic Steel
Microalloying
Microstructural Evolution
Microstructure
Niti Coating
Plasma Transferred Arc Hardfacing
porous materials
Realized Kernels
Rock Microstructure
service life design
structural materials
waste inclusion

Synthetic cell technology

Abiogenesis
Actin Cytoskeleton
Artificial Cells
bioengineering
Biomanufacturing
Build-a-cell
Cell transport
Cortactin
CRISPR technology
Cytoskeletal Proteins
Cytoskeleton
Metabolic Engineering
molecular machinery
Origin of Life
Podosomes
Septins
Single-molecule protein sequencing
Synthetic Biology

synthetic cell
Tropomodulin
Thin films and coatings
Absorbers (materials)
amorphous film
Amorphous Semiconductors
atomic layer deposition
Atomic Layer Epitaxy
Cadmium Sulfide Solar Cells
Chemical Vapour Deposition
coating technologies
coating technology
coil coat
electro coat
electrostatic spraying
ferroelctric film
Ferroelectric Films
Film Preparation
Gallium Oxides
Hafnium Oxides
Indium Sulfide
Ito Glass
metallic film
nanocomposite film
Nanocomposite Films
Nanosheet
optical film
Optical Films
Oxide Semiconductors
powder coating
Pulsed laser deposition
Semiconducting Selenium Compounds
Semiconductor Doped Polymers
Sol-gel
Sol-gel Process
spray coating
Thin Film Circuits
Thin Film Equation
Thin Film Flow
Thin Film Lithium Ion Batteries
Thin Film Solar Cells
Thin Film Transistors
Thin Films
waterborne coating

Bijlage 3: Resultaten van de tweede validatie door WBSO-experts

Tabel B3.1. Het aantal projecten gevonden met de 'concepts or keywords' van Elsevier (E), met de zoektermen van de WBSO-adviseurs (W) en met de ngrams uit de WBSO-projectbeschrijvingen (N)

Gevonden met / Sleuteltechnologie	alleen met E, niet met N of W	alleen met W, niet met E of N	alleen met N, niet met E of W	met E en W, niet met N	met E en N, niet met W	met N en W, niet met E	met E, N en W	totaal aantal gevonden projecten
Artifical intelligence (incl. machine and deep learning)	40	2.457	10.109	560	91	1.919	1624	16.800
Genomics/proteomics/metabolomics/ glycomics/X-omics	13	1.000	4.163	19	62	313	41	5.611
Imaging technologies	38	5.458	9.810	39	38	2.195	64	17.642
Thin films and coatings	0	42	26.582	0	13	20	0	26.657
Additive manufacturing/3D printing	0	0	444	0	232	999	45	1.720
Big data and data analytics	0	0	632	0	1.898	1.299	201	4.030
Energy storage materials	0	0	0	0	0	4.433	200	4.633
Gene editing/precise genetic engineering	11	736	3.382	0	38	131	23	4.321
High frequency and mixed signal technologies	0	0	23.406	0	283	1.399	31	25.119

Nota bene: bovenstaande categorieën zijn wederzijds uitsluitend en tellen op tot het totaal aantal gevonden projecten.

Bijlage 4: Conclusies van de expertsessie

Expertsessie Sleuteltechnologieen WBSO op basis van text mining
Locatie: RVO.nl, Utrecht, 25 februari 2019, 14.00-16.00 uur

Doel
De expertsessie is voor de projectgroep geslaagd als er feedback wordt gegeven op de gekozen methode en de uitgevoerde analysestappen.
Zijn er verbetermogelijkheden in de aanpak en analysestappen? Wat zijn mogelijke alternatieve wegen die perspectief bieden op betere resultaten?

Programma

Welkom/voorstelronde (Pieter de Bruijn)	14:00-14:10
Thematische analyses WBSO – huidige methode (Koen Septer)	14:10-14:20
Elsevier text mining tools (Jeroen Geertzen)	14:20-14:30
Aanpak en conceptresultaten haalbaarheidsstudie (Edwin Horlings)	14:30-15:00
Feedback en discussie	15:00-15:50
Conclusie en afsluiting	15:50-16:00

Conclusie
In relatie tot het doel van onze haalbaarheidsstudie – verkennen in hoeverre de aanpak van het eerdere Elsevier onderzoek op basis van wetenschappelijke publicaties toepasbaar is om de inzet op sleuteltechnologieën vanuit de WBSO in kaart te brengen – is tijdens de sessie het volgende beeld ontstaan:

1. Toepassing van de Elsevier methode en tool op de WBSO behelst duidelijk meer dan slechts een spreekwoordelijke 'druk op de knop'
   - de key words zijn gesteld in het Engels terwijl de projectomschrijvingen WBSO voor het overgrote deel (94%) in het Nederlands zijn geschreven
   - de Elsevier tool (normalisatie van projectomschrijvingen) werkt (beter) op basis van Engelse en werkt niet c.q. minder op Nederlandse taal (Nederlandstalige module is bij dit onderzoek c.q. light versie 'uitgezet')
   - de onderzoeksaanpak van Elsevier heeft zijn waarde goed bewezen in context wetenschappelijke publicaties; bij projectomschrijvingen WBSO is sprake van meer variatie in taalgebruik dan in wetenschappelijke tijdschriften het geval is
2. Ondanks dat de Elsevier methode niet puur een druk op een knop is, levert onze studie toch aardige eerste inzichten op
   - de tool levert bruikbare resultaten op; wel valt op dat de resultaten flink verschillen tussen Nederlandstalige en Engelstalige projectvoorstellen (in eerste run EN 50% binnen KET; NL 15% binnen KET)
   - een eerste run levert al een aardige indicatie op; o.b.v. van check door WBSO-adviseurs blijkt evenwel dat we er nog niet zijn (op to do lijst voor komende weken staat nog een feedback loop op de agenda alvorens een tweede run uit te voeren)
3. De Elsevier methodiek lijkt in grote lijnen sterk op de methode die eerder door de WBSO werd toegepast, maar er zijn ook verschillen
   - beide methoden kijken in hoeverre vooraf bepaalde key words terugkomen in teksten
   - de gewogen trefwoordenanalyse discrimineert tussen trefwoorden (aan sommige woorden wordt meer waarde toegekend dan aan andere woorden)
   - bij gewogen trefwoordenanalyse worden varianten van een en hetzelfde woord handmatig toegevoegd aan de trefwoordenlijst; bij Elsevier tool is

dit voor Engelse teksten niet nodig (normalisatie van projectomschrijvingen)

4. Voor het vervolg geven experts aan dat een bottom-up strategie kansen biedt
   Bij een bottom-up strategie start je niet vanuit vooraf bepaalde trefwoorden, maar start je door WBSO experts naar een aantal typische projecten binnen een bepaalde KET te vragen; vervolgens laat je de software los op de projectomschrijvingen van deze specifieke set projectomschrijvingen (die je op deze manier gebruikt als een trainingsset in zgn. supervised machine learning)

Bijlage 5: Lijst met zoektermen die door de WBSO-adviseurs zijn gevalideerd

Aan de WBSO-adviseurs die bij dit onderzoek betrokken waren, is een lijst met zoektermen voorgelegd die betrekking hebben op één specifieke sleuteltechnologie. Deze lijst bestond is (a) de zoektermen van Elsevier, (b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde, en (c) termen en concepten die voorkomen in de door de experts gevalideerde projecten.

We hebben de WBSO-adviseurs gevraagd om iedere zoekterm een score te geven, namelijk:
0 = Niet relevant, geen goede zoekterm.
1 = Relevant maar te generiek.
2 = Relevant en specifiek genoeg, maar op zichzelf onvoldoende.
3 = Relevant en specifiek, op zichzelf voldoende om de sleuteltechnologie te vinden.

Sleuteltechnologie: Artifical intelligence (incl. machine and deep learning)

(a) de zoektermen van Elsevier
Ant Colony Optimization (0)
artificial intelligence (2)
Autonomous decision making (2)
Autonomous systems (2)
Back-propagation Neural Network (2)
Bam Neural Networks (2)
Boltzmann Machine (0)
Bp Neural Network (1)
Causal interference (0)
Cohen-grossberg Neural Networks (1)
Computational Creativity (0)
Co-training (0)
Deep Learning (3)
Delayed Neural Networks (2)
Echo State Network (0)
Elman Neural Network (0)
Extreme Learning Machine (0)
Genetics-based Machine Learning (2)
Machine learning (2)
Neural networks (2)
Neuroevolution (0)
Probabilistic Neural Network (2)
Pulse Coupled Neural Network (1)
Radial Basis Function Neural Network (1)
Recurrent Neural Networks (2)
Reinforcement Learning (2)
Semi-supervised Learning (2)
Sentiment Classification (2)
Spiking Neural Networks (1)
Stochastic Neural Networks (2)
Supervised Learning (2)
Supervised Machine Learning (2)
Swarm Intelligence (1)
Turing Test (2)
Unsupervised Machine Learning (2)
Wavelet Neural Network (2)

(b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde
voorspellen (2)
predictive (3)
zelflerend (3)
learning (2)
neurale netwerken (3)
CNN (3)
genetische algoritmen (3)
R (als programmeertaal) (2)
Python (idem) (2)
tensor (2)
TensorFlow (2)

(c) termen en concepten die voorkomen in de door de experts gevalideerde projecten
neural networks (2)
neural network (2)
neuraal netwerk (2)
neurale netwerken (2)
natural language processing (2)
NLP (2)
natural language (2)
natuurlijke taal (2)
deep learning (3)
machine learning (3)
unsupervised machine learning (3)
unsupervised learning (2)
machine learning platform (2)
machine learning platforms (2)
text normalization (1)
text normalisation (1)
tekstnormalisatie (1)
social media (1)
artificial intelligence (2)
AI (2)
kunstmatige intelligentie (2)
performance verbetering (0)
service platform (0)
algoritme (1)
algoritmen (1)
algorithm (1)
algorithms (1)

fuzzy logic (1)
PoS tags (0)
speech recognition (1)
spraakherkenning (1)
input text (0)
text modules (0)
tekstmodules (0)
information processing (0)
informatieverwerking (0)
taal (0)
language (0)
database (0)
voorspelling (2)
voorspellingen (2)
prediction (2)
predictions (2)
bot (2)
bots (2)
computationele fysische simulaties (1)
computationele fysische simulatie (1)
decision tree (2)
point cloud (0)
patronen herkennen (0)
complex model (0)
complex models (0)
foto (0)
API (0)
cloud (0)

Sleuteltechnologie: Genomics/proteomics/metabolomics/ glycomics/X-omics

(a) de zoektermen van Elsevier
bioinformatics (2)
Comparative Genomics (2)
epigenomics (2)
Exome (2)
Functional Genomics (2)
genome analysis (2)
Genome Assembly (2)
genomic engineering (2)
Genomic Structural Variation (2)
genomics (2)
Glycomics (2)
High-throughput nucleotide Sequencing (2)
Metabolome (2)
metabolome analysis (2)
Metabolomics (2)
Metagenomics (2)
Molecular Sequence Annotation (2)
National Human Genome Research Institute (u.s.) (0)
next generation sequencing (2)
proteome analysis (2)
Proteomics (2)
RNA sequencing (2)
Transcriptomics (2)
Two-dimensional Difference Gel Electrophoresis (2)
Unigenes (2)
whole exome sequencing (2)

(b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde
sequencing (2)
bioinformatica (2)
genetic engineering (2)
enzymtechnologie (2)
proteomics (2)
modificeren van cellen (2)
muteren van genen (2)
gentherapie (3)
enzymtechnologie (2)
kit ontwikkeling (3)
analytisch diagnostiek (3)
metaboliet analyse (3)
farmacokinetiek (3)
PK (3)
plasmide (2)
vector (2)
stamcel (3)
in situ (2)
histochemie (2)

(c) termen en concepten die voorkomen in de door de experts gevalideerde projecten
Next Generation Sequencing (2)
NGS (2)
whole genome sequencing (2)
RNA binding proteins (2)
RNA binding protein (2)
broad spectrum (1)
cross-protective (1)
genenpool (2)
rDNA sequencing (2)
RNA sequencing (2)
cell line generation (2)
cell line (2)
cell lines (2)
signaal transductie paden (2)
signaal transductie (2)
genetische koppelingen (2)
genetische koppeling (2)
cross-protective inactivated vaccine (2)
inactivated vaccine (2)
veredelingstechnieken (2)
veredelingstechniek (2)
veredeling (2)
generic modifiers (2)
generic modifier (2)
immune cell (2)
immune cells (2)
immune cell count (2)
immune cell counts (2)
hybridisaties (2)
hybridisatie (2)
terugkruisingen (2)
terugkruising (2)
kruisingsveredeling (2)
kruisingen (2)
kruising (2)
genetic drag (2)
RNA binding (2)
binding proteins (2)
bacteriologische diagnostiek (2)
conserved domains (1)
cell generation (2)
cell kinetics (2)
DNA informatie (2)
DNA-methylatie (2)
gene expression (2)
host restriction (2)
encode specificity (2)
genen identificeren (1)

genetische variatie (2)
H9N2 genotypes (2)
genotypes (2)
interspecifieke hybridisaties (2)
intraspecifieke hybridisaties (2)
invasive behaviour (0)
reverse phenotyping (2)
targeted sequencing (2)
breeding (1)
resistentie (1)
markers (1)
marker (1)
merkers (1)
merker (1)
biomarkers (2)
biomarker (2)
genome (2)
genoom (2)
bioinformatica (2)
bioinformatics (2)
ras (0)
rassen (0)
genomics (2)
genomic (2)
genetisch (2)
genetische (2)

Sleuteltechnologie: Imaging technologies

(a) de zoektermen van Elsevier
Bolometers (2)
Cardiac-gated Imaging Techniques (3)
Cardiac-gated Single-photon Emission Computer-assisted Tomography (3)
Computer Assisted Image Interpretation (3)
Computer Assisted Radiographic Image Interpretation (3)
Cone-beam Computed Tomography (3)
Diffuse Optical Tomography (3)
Diffusion Tensor Imaging (3)
Discrete Tomography (3)
Dual Energy Scanned Projection Radiography (3)
Electric Impedance Tomography (3)
Electrical Capacitance Tomography (3)
Electrical Impedance Tomography (3)
Electrical Resistivity Tomography (3)
Electron Microscope Tomography (3)
far-infared imaging (3)
Fiducial Markers (3)
Filtered Backprojection (2)
Four-dimensional Computed Tomography (3)
Free-form Deformation (3)
Geometric Tomography (3)
Image Encryption (3)
Image Inpainting (3)
image processing (3)
image reconstruction (3)
Image Registration (3)
image-guided intervention (3)
Image-guided Radiotherapy (3)
Imaging technology (3)
Infrared Detectors (2)
infrared imaging (3)
Lens Antennas (2)
Medical Image Processing (3)
Medical imaging (3)
Micro-ct (3)
Microresonators (3)
Microwave Resonators (3)
molecular imaging (3)
Multidetector Computed Tomography (3)
Multimodal Imaging (3)
Noise Temperature (2)
Optical Coupling (2)
Optical imaging (3)
Optical Tomography (3)
Phase Contrast (2)
Photoacoustic Tomography (3)
Positron Emission Tomography (3)
Radio Astronomy (3)
Radio imaging (3)
Remote Sensing Image (3)
Respiratory-gated Imaging Techniques (3)
Retinal Photoreceptor Cell Inner Segment (2)
Seismic Refraction Tomography (3)
Seismic Tomography (3)
Stereoscopy (3)
Submillimeter Waves (2)
Super-resolution (2)
Terahertz imaging (3)
Terahertz Radiation (2)
Terahertz Waves (2)
tomography (3)
Ultrasound Image (3)
Whole Body Imaging (3)
X-ray imaging (3)
X-ray Microtomography (3)

(b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde
beeldverwerking (3)
Autonomous Driving (2)
AD (2)
radar (3)
camera (3)
lidar (3)
litho process (3)
imaging (3)
beeldherkenning (3)
beeldherkenningssysteem (3)
beeldherkenningssystemen (3)
lithography (3)
belicht (1)
computer vision (3)
vision (3)
slimme camera (2)

(c) termen en concepten die voorkomen in de door de experts gevalideerde projecten
image processing algorithm (3)
image processing algorithms (3)
image processing technique (3)
image processing techniques (3)
image processing techniek (3)
image processing technieken (3)
image processing software (3)
multiscale image processing (3)
image processing (3)
image acquisition (3)
imaging technology (3)

imaging technologies (3)
3D (1)
optical imaging (3)
MRI (3)
functional MRI (3)
fMRI (3)
bone MRI (3)
medical image (3)
medical imaging (3)
x-ray (3)
x-ray angiografie (3)
electron microscope (2)
electron microscopes (2)
microscope settings (2)
diffusion tensor (3)
Hough transforms (3)
electron beam (2)
electron beams (2)
electron detector (2)
electron detectors (2)
dynamisch programmeren (0)
visual studio (0)
object-georiënteerde ontwikkelmethode (0)
control software (0)
low-level segmentatie library (0)
canny edge detection (3)
optimale contour overlay (3)
pattern recognition (2)
patroonherkenning (2)
geknipte pillen (0)
halve pillen (0)
video analytics (2)
phase plate (3)
fantoom experimenten (0)
object matching (2)
.NET framework (0)
.NET tools (0)
dynamische thresholding (1)
spectraal camera (3)
detectie (1)
detector (2)
detectors (2)
CT (3)

Sleuteltechnologie: Thin films and coatings

(a) de zoektermen van Elsevier
Absorbers (materials) (2)
amorphous film (3)
Amorphous Semiconductors (2)
atomic layer deposition (3)
Atomic Layer Epitaxy (3)
Cadmium Sulfide Solar Cells (3)
Chemical Vapour Deposition (3)
coating technologies (3)
coating technology (3)
coil coat (3)
electro coat (3)
electrostatic spraying (3)
ferroelctric film (3)
Ferroelectric Films (3)
Film Preparation (3)
Gallium Oxides (3)
Hafnium Oxides (3)
Indium Sulfide (3)
Ito Glass (3)
metallic film (3)
nanocomposite film (3)
Nanocomposite Films (3)
Nanosheet (3)
optical film (3)
Optical Films (3)
Oxide Semiconductors (3)
powder coating (3)
Pulsed laser deposition (3)
Semiconducting Selenium Compounds (3)
Semiconductor Doped Polymers (3)
Sol-gel (2)
Sol-gel Process (2)
spray coating (3)
Thin Film Circuits (3)
Thin Film Equation (3)
Thin Film Flow (3)
Thin Film Lithium Ion Batteries (3)
Thin Film Solar Cells (3)
Thin Film Transistors (3)
Thin Films (3)
waterborne coatin (3)

(b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde
ferroelectric film (3)
Langmuir-Blodgett (3)
Monolayer (3)
Multilayer (3)
Electroplating (3)
Electrodeposition (3)
Protective coating (3)
Anti-bacterial coating (3)
Optical coating (3)
Sputtering (3)
Spin coating (3)
Dip coating (3)
Molecular Beam Epitaxy (3)
Antibacterial coating (3)
Functional coating (3)
Barrier (3)
Liquid crystalline (3)

(c) termen en concepten die voorkomen in de door de experts gevalideerde projecten
Atomic Layer Deposition (3)
ALD (3)
chemical vapor deposition (3)
chemical vapour deposition (3)
chemical deposition (3)
meerlaags (3)
niet-dekkende (3)
niet-dekkend (3)
verfproducten (3)
verfproduct (3)
buisoven reducerend gas (?)
sol-gel (2)
sol-gel route (2)
sol-gel routes (2)
carbon-fiber reinforced plastic (0)
CFRP (0)
coating (3)
coatings (3)
metallische deeltjes (1)
gevormde lagen kristallijn (3)
kristallijn (1)
noble metal catalysts (1)

noble metal catalyst (1)
infrarood licht (1)
IR-licht (1)
IR-licht reflecteren (1)
powder coating resins (3)
powder coating resin (3)
pigment coating (3)
pigment coatings (3)
powder coating (3)
powder coatings (3)
resins (1)
resin (1)
thermische prikkels (0)
zonne-energie (0)
solar cells (1)
solar cell (1)
reflecteren (1)
reflectie (1)
surface modification (2)
anorganische laag (2)
nail polish (2)
nagellak (2)
anorganische coating (3)
coating material (3)
coating technologies (3)
plasma deposition (3)
porositeit (1)
radiation heating (1)
laag (1)
lagen (1)
adhesion (1)
nanodeeltjes (1)
cold end (1)
field joint (?)
fotokathode laag (3)
folie (1)

Sleuteltechnologie: Additive manufacturing/3D printing

(a) de zoektermen van Elsevier
3D model (0)
3D model retrieval (1)
3d Modeling (1)
3D modelling (1)
3d Printers (3)
3D printing (3)
additive manufacturing (3)
binder jetting (3)
Biofabrication (3)
Bioprinting (3)
CAM (0)
Cam-clay Model (1)
Computer Aided Manufacturing (3)
Digital twinning (0)
directed energy deposition (3)
Electron Beam Melting (3)
filament extrusion (3)
fused deposition modeling (3)
Layered Manufacturing (3)
Manufacturing Technology (0)
material extrusion (3)
material jetting (2)
Medical splints (0)
powder bed fusion (3)
Printing Machinery (3)
selective laser melting (3)
Selective Laser Sintering (3)
Selective laser sintering (3)
sheet lamination (3)
Splints (medical) (0)
Stereolithography (3)
vat photopolymerization (3)
materiaal extrusie (3)
sheet lamineren (3)
laser sinteren (3)
laser smelten (3)
poeder bed (3)
printer machine (3)

(b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde
filament (2)
printbare stof (2)
3D printen (3)
3D printer (3)
printkop (3)
foodprinting (3)
technieken voor 3D-printen op een rijtje: (0)
Stereolithografie (3)
SLA (3)
Laser sinteren (3)
sinteren (3)
Fused deposition modelling (3)
FDM (3)
Material jetting (3)
Photopolymer jetting (3)
Binder jetting (3)
Lasersmelten (3)
Elektronenstraal smelten (3)
3D (laser)cladden (3)
Selective Laser Sintering (3)
SLS (3)
3D metal printing (3)
metal printing (3)
metaal printen (3)
Computer Aided Manufacturing (3)
Cold metal spray (3)
Foodprinting (3)

(c) termen en concepten die voorkomen in de door de experts gevalideerde projecten
piezo actuated micro hopper nozzle (3)
piezo actuated (3)
nozzle design (1)
additive manufacting (3)
food printing (3)
automated production (0)
printing technologie (3)
printing technologieën (3)
printing technology (3)
printed circuit (3)
powder flow (2)
laser melting (3)
3D printer (3)
3D printers (3)
3D-printer (3)
3D-printers (3)
3D printen (3)
3D-printen (3)
3D printing (3)
3D model (0)
3D models (0)
material input (0)

materiaaleigenschappen (0)
plastic (0)
industrieel afval (0)
stepper drivers (0)
stepper driver (0)
Fused Deposition Modeling (3)
Selective Laser Melting (3)
injection moulding matrijs (2)
injection moulding matrijzen (2)
injection moulding (2)
matrijs (0)
matrijzen (0)
filament (3)
filamenten (3)

Sleuteltechnologie: Big data and data analytics

(a) de zoektermen van Elsevier
Apriori Algorithm (0)
Automatic Image Annotation (1)
Big Data (2)
Data accuracy (1)
Data confidentiality (1)
Data fairness (1)
Data Mining (2)
data science (2)
data stewardship (1)
Data transparency (1)
Deep Learning (1)
distributed sensor data (2)
Efficient deep learning (1)
Findable data (1)
Frequent Pattern Mining (2)
heterogeneous data (2)
image analysis (2)
Image Annotation (1)
information retrieval (0)
learning algorithms (1)
machine learning (1)
Mapreduce (2)
Opinion Mining (2)
Pattern Mining (2)
Privacy Preserving Data Mining (1)
Process Mining (2)
radiomics (1)
Responsible data (1)
Re-usable storage of data (0)
Sentiment Analysis (2)
Text Analysis (2)
Unsupervised Machine Learning (0)
Visual Analytics (1)

(b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde
big data (1)
regressie (2)
correlatie (2)
A-B-testing (0)
Mining (2)
No-SQL (1)
indexering (2)
sentiment (1)

(c) termen en concepten die voorkomen in de door de experts gevalideerde projecten
machine learning techniques (0)
machine learning technique (0)
machine learning (0)
content delivery (1)
deep learning (0)
algoritme (0)
algorithm (0)
algoritmes (0)
algorithms (0)
deep learning (2)
learning algoritme (1)
learning algorithm (1)
learning algoritmes (1)
learning algorithms (1)
Jenkins automation server (0)
Jenkins server (0)
multichannel content delivery (1)
multi channel content delivery (1)
content delivery (1)
Natural Language Processing (1)
regular expressions (0)
regular expression (0)
analytics engine (0)
parsing (0)
social media platforms (0)
social media platform (0)
social media (0)
3D-vision motion tracking (0)
neurale netwerken (1)
neuraal netwerk (1)
patronen herkennen (1)
motion tracking (1)
semantische relaties (1)
PHP (0)
API (0)

Sleuteltechnologie: Energy storage materials

(a) de zoektermen van Elsevier
Alkaline Batteries (3)
Alkaline Battery (3)
batteries (2)
Battery (2)
Battery Chargers (2)
Battery Disposal (2)
Battery Electric Vehicles (2)
Battery Management Systems (2)
Button-cell Battery (2)
CAES (2)
CAES diabatic (0)
Catholytes (0)
Charging (batteries) (2)
Compressed Air Energy Storage (2)
Compressed Air Energy Storage (CAES) (2)
DC compressor (0)
direct current compressor (0)
Electric Battery (2)
Electric Current Collectors (2)
Electrochemical Cells (2)
electrochemical condensator (2)
electrochemical storage (3)
electromechanical storage (3)
Energy Storage (3)
energy storage material (3)
Flow Batteries (2)
flywheel energy storage (3)
Flywheel Propulsion (2)
Fuel Cells (2)

Inductive Power Transmission (0)
Ion Storage (1)
Iron Phosphates (0)
kinetic energy storage (1)
Lead Acid Batteries (3)
Lithium Batteries (3)
Lithium Compounds (1)
Lithium Deposits (1)
Lithium Sulfur Batteries (3)
Lithium-ion Batteries (3)
Metal Air Batteries (3)
Miniature Batteries (3)
Nickel Metal Hydride Batteries (3)
Nuclear Batteries (2)
Plug-in Electric Vehicles (1)
Power-to-power (P2P) (1)
Primary Batteries (2)
Pulse Charging (2)
pumped hydroelectric storage (PHS) (2)
Shell Anodes (0)
Silicon Batteries (3)
Superconducting Magnetic Energy Storage (SMES) (2)
Thermal Batteries (2)
Thin Film Lithium Ion Batteries (3)
Zinc-bromide Batteries (3)
Zinc-oxygen Batteries (3)

(b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde
Geen aanvullende zoektermen nodig. Alleen in de titel zoeken, anders komen er allemaal andere projecten naar voren die niets met de ontwikkeling van sleuteltechnologie "Energy storage materials" te maken hebben. Zoektermen zoals "batterijen" en "battery" zijn allemaal op 3 gezet als ze in de titel voorkomen.

(c) termen en concepten die voorkomen in de door de experts gevalideerde projecten
lithium battery packs (3)
lithium battery (3)
lithium batteries (3)
lithium batterij (3)
lithium batterijen (3)
lithium (0)
Battery Management System (2)
BMS (0)
batterij (3)
batterijen (3)
battery pack (3)
battery packs (3)
phase change materials (3)
phase change material (3)
energy storage (3)
energie opslag (3)
energieopslag (3)
batterij technologie (2)
operationele betrouwbaarheid (0)
opgewekte energie (0)
geleidingsplaat (0)
operationele levensduur (0)
stroomdichtheden (0)
stroomdichtheid (0)
voltages (0)
voltage (0)
cathode material (1)
accu (2)
pack (1)
packs (1)
capaciteit (0)
opslagcapaciteit (1)
laadstroom (2)

Sleuteltechnologie: Gene editing/precise genetic engineering

(a) de zoektermen van Elsevier
Cas9 (3)
Clustered Regularly Interspaced Short Palindromic Repeats (2)
Cofactor Engineering (1)
Comparative Genomics (3)
CRISPR (3)
Crispr-associated Proteins (3)
Crispr-cas Systems (3)
Dna End-joining Repair (2)
gene editing (3)
gene therapy (3)
gene transfer (3)
gene transfer techniques (3)
Genetic engineering (3)
Genetic Modification (3)
genome analysis (2)
Genome Assembly (2)
Genomic Structural Variation (2)
Guide Rna (2)
Inverted Repeat Sequences (2)
Lentivirus (1)
Metabolic Engineering (0)
Molecular Farming (2)
Molecular Sequence Annotation (2)
Nuclear Transformations (2)
Targeted Gene Repair (3)

(b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde
sequencing (2)
bioinformatica (2)
genetic engineering (3)
enzymtechnologie (0)
proteomics (0)
modificeren van cellen (2)
muteren van genen (3)
gentherapie (3)
enzymtechnologie (0)
kit ontwikkeling (1)
analytisch diagnostiek (1)
metaboliet analyse (0)
farmacokinetiek (0)
PK (0)
plasmide (1)
vector (1)
stamcel (1)
in situ (1)
histochemie (0)

(c) termen en concepten die voorkomen in de door de experts gevalideerde projecten
DNA editing technology (3)
vererving (1)
genetische koppelingen (2)
genetische koppeling (2)

gene therapy product (2)
stem cell lines (1)
cell-based AAV production platform (1)
cell-based AAV production platforms (1)
CRISPR (3)
kinase (1)
gene transfer (2)
viral variant (1)
viral variants (1)
cell line (1)
cell lines (1)
cell culture (1)
cell cultures (1)
resistentieveredeling (2)
veredeling (2)
DNA editing (3)
DNA modulation (3)
gene therapy (3)
DNA oligonucleotide targeting (2)
DNA repair systems (2)
DNA repair system (2)
DNA repair (2)
RNA repair (2)
eiwit producerende genen (1)
editing efficiency (2)
hybriden (2)
kruisingsouders (2)
inkruisen (2)
genetische bronnen (2)
genetische bron (2)
genetische analyse (2)
genetische analyses (2)
interspecifieke hybridisaties (1)
intraspecifieke hybridisaties (1)
hybridisaties (1)
terugkruisingen (1)
terugkruising (1)
marker assisted selection (1)
modulation technology (0)
genetische variatie (2)
hybride rassen (1)
protein expression (1)
editing technology (2)
stem cell (1)
stem cells (1)
DNA technology (1)
gene editing (3)
veredelingstechnieken (2)
veredelingstechniek (2)
veredeling (2)
genetische variatie (2)
plant breeding (2)
gene transfer (3)
genome editing (3)
gene expression (2)
genetic drag (2)
AAV (1)
flowcytometrische analyses (1)
flowcytometrische analyse (1)
genetic disorders (2)
genetic disorder (2)
genome testing (2)
gene encoding (2)
ras (1)
rassen (1)
onderstammen (1)
CRISPR/CAS (3)
resistenties (0)
resistentie (0)
mutations (1)
mutation (1)
mutaties (1)
mutatie (1)
cell culture assays (1)
scheutregeneratie (1)
protein expression (1)

Sleuteltechnologie: High frequency and mixed signal technologies

(a) de zoektermen van Elsevier
5G (2)
5G mobile communication systems (2)
Antenna feeders (3)
Antenna phased arrays (3)
aperture arrays (3)
Beamforming (3)
Bolometers (3)
CMOS (3)
Coplanar Waveguides (3)
Cryogenics (0)
Demodulation (2)
Detection (1)
Electromagnetic Wave Polarization (2)
Focal Plane Arrays (3)
Frequency Bands (2)
Frequency Modulation (2)
Frequency Response (2)
Frequency Selective Surfaces (2)
HEB technology (0)
Heterodyne (3)
Heterodyne technology (3)
High Frequency (3)
high frequency signal technology (3)
hot electron bolometer (3)
Infrared Detectors (3)
Lens Antennas (3)
low noise amplifiers (3)
Low power converter (3)
Massive MIMO (3)
Metamaterials (1)
Microbolometer (3)
Microwave Devices (3)
Microwave KID technology (3)
Microwave Resonators (3)
Mixed signal integrated circuit (3)
mixed signal technology (3)
MOSFET (3)
Multiple-input multiple-output (mimo) Systems (3)
Noise Temperature (3)
Phase Modulation (2)
phased array (3)
phased array feed (3)
Q Factor (3)
Radio Astronomy (3)
Responsivity (1)
Sidebands (3)
Submillimeter (2)
Submillimeter Waves (3)
superconducting detectors (3)
Superconducting Devices (3)
Superconducting Resonators (3)
Surface Plasmon (3)
Terahertz (3)
Terahertz Imaging (3)

Terahertz Radiation (3)
Terahertz Spectroscopy (3)
Terahertz Wave Detectors (3)
Terahertz Waves (3)
TES technology (3)
Thermal Energy (3)
THz sensing (3)
Titanium Nitride (2)
transition edge sensors (2)
Tunable (1)
Tunnel Junctions (3)
Ultra-wideband (uwb) (3)
Waveguides (2)

(b) termen en concepten uit de feedback van de betreffende WBSO-adviseur uit de validatieronde
radar (2)
IC ontwikkeling (3)
GSM (1)
ASIP (3)
ASIPs (3)
System on a chip (3)
SOC (3)
5G (2)
Single chip DECT (3)
Multi Level Modulatie (2)
300MHz (3)
Antenna Sub-Systeem (3)
antenna (3)
hoge bandbreedte (2)
Smart Antenna (3)
GHz (3)

(c) termen en concepten die voorkomen in de door de experts gevalideerde projecten
Low Noise Amplifier (3)
programmeerbare elektronische componenten (2)
programmeerbare electronische componenten (2)
phased array transducer (3)
phased array (3)
fiber optic interrogation (3)
embedded software (1)
opgewekte elektriciteit (1)
opgewekte electriciteit (1)
mobiele telefoon (2)
Internet of Things (2)
IOT (2)
parametriseerbaar netwerk (2)
RGB camera (2)
5G (2)
SD (0)
netwerk (1)
netwerken (1)
sensoren (3)
sensor (3)
draadloos (2)
draadloze (2)
signaal (1)
detectie (1)
detection (1)

Bijlage 6: Source van de Python tools gebruikt in deze studie

Taalherkenning

```
#TOKENIZER AND LANGUAGE DETECTOR
#Version 1
#Edwin Horlings
#CBS, November 2018

#This script detects the language of the full-text data automatically using detect_langs,
selectin between
#English and Dutch. The table that is produced is input for tokenization, making stopword lists
and finding ngrams.

#Modules

import sqlite3
from sqlite3 import Error
from langdetect import detect, detect_langs
import time

def English_Or_Dutch(string):
    res = detect_langs(string)
    for item in res:
        if len(string) > 0 and (item.lang == "nl" or item.lang == "en"):
            return item.lang
    return None

#Main

def main():

    #Global variables

    global conn, cursor

    #Connect to database

    conn = sqlite3.connect("C:\WERK\CBS\RVO\data\wbso.db")
    cursor = conn.cursor()

    #Create output tables

    sql = """DROP TABLE IF EXISTS wbso_languages"""
    cursor.execute(sql)
    sql = """CREATE TABLE wbso_languages (project_id INTEGER, language TEXT)"""
    cursor.execute(sql)

    #Read full-text data from MySQL database

    query = """SELECT project_id, project_name, fp_programme, description FROM
wbso_master"""
    cursor.execute(query)
    result = cursor.fetchall()

    num_records = len(result)
    x=0
    while x < num_records:
        this_record = list(result[x])
        project = this_record[0]

        language_row = []
        language = English_Or_Dutch(this_record[3])      #[1]=titel, [2]=fp_programme,
[3]=description
```

```
        language_values = (project, language)
        language_row.append(language_values)

        sql = """INSERT INTO wbso_languages (project_id, language) VALUES (?, ?)"""
        cursor.executemany(sql,language_row)
        conn.commit()
        print(x)

        x += 1

    sql = """CREATE INDEX ix_project_id ON wbso_languages (project_id)"""
    cursor.execute(sql);

    cursor.close()
    conn.close()

    end_time = time.process_time()
    print(end_time)

if __name__ == '__main__':
    main()
```

Ngrams uit de tekst van projectomschrijvingen halen

```
#NGRAM FINDER
#Version 1
#Edwin Horlings
#CBS, February 2019

#This script determines the frequency of ngrams (single words, bigrams and trigrams) in full-text information.
#It first tokenizes the text, which produces a table with all individual tokens in the text, and then identifies
#unigrams, bigrams, and trigrams, controlling for the words on a stopword list (Dutch and English).

#Modules

import sqlite3
from sqlite3 import Error
from nltk.corpus import stopwords
import nltk.tokenize, re, pprint
from nltk.tokenize import word_tokenize, punkt, sent_tokenize, RegexpTokenizer
import nltk.tag.util
from nltk.tag.util import tuple2str
from nltk.collocations import *
from langdetect import detect, detect_langs
import time

#Functions

def import_text_file(location,charmap):

    file = open(location, 'r', encoding=charmap)     #Load text file
    rows = file.read().splitlines()                  #read without \n
    file.close()

    return rows

def store_ngram(list_of_values):

    num_ngrams = len(list_of_values)

    selected_text_list = []
    x=0
```

```
    while x < num_ngrams:
        this_record = list_of_values[x]

        #Fields to merge into one blob variable
        selected_fields = (str(this_record[0]), str(this_record[1]), this_record[2], this_record[3],
str(this_record[4]))
        text_field = '\t'.join(selected_fields)
        out_record = (text_field, '\n')
        text_to_write = ''.join(out_record)
        selected_text_list.append(text_to_write)
        x += 1

    selected_text = ''.join(selected_text_list)
    out_file.write(selected_text)

def english_or_dutch(string):

    res = detect_langs(string)
    for item in res:
        if len(string) > 0 and (item.lang == "nl" or item.lang == "en"):
            return item.lang
        else:
            return "xx"

def tokenize_text():

    #Create output tables in database
    sql = """DROP TABLE IF EXISTS tokens"""
    cursor.execute(sql)
    sql = """CREATE TABLE tokens (project_id INTEGER, token TEXT, position INTEGER,
language TEXT)"""
    cursor.execute(sql)

    #Extract tokens
    x=0
    while x < num_records:
        this_record = text_database[x]
        project = this_record[0]
        language = this_record[2]

        tokens = word_tokenize(this_record[1])

        token_row = []
        num_tokens = len(tokens)
        y = 0
        while y < num_tokens:
            token_pos = y+1
            if len(tokens[y]) <= 40:  #max length of 40 weeds out hyperlinks
                if language == 'en':
                    if tokens[y].lower() not in stopwords_en:
                        token_values = (project, tokens[y], token_pos, language)
                        token_row.append(token_values)
                elif language == 'nl':
                    if tokens[y].lower() not in stopwords_nl:
                        token_values = (project, tokens[y], token_pos, language)
                        token_row.append(token_values)

            y += 1

        sql = """INSERT INTO tokens (project_id, token, position, language) VALUES (?, ?, ?,
?)"""
        cursor.executemany(sql,token_row)

        conn.commit()

        x += 1

def find_unigrams(min_occurrence):
```

```
    ngram_type = 1

    #Tokenize
    print('--Tokenizing')
    tokenize_text()

    #Create output tables in database
    sql = """DROP TABLE IF EXISTS token_frequency"""
    cursor.execute(sql)
    sql = """CREATE TABLE token_frequency (project_id INTEGER, token TEXT, language TEXT,
frequency INTEGER)"""
    cursor.execute(sql)

    #Insert token frequencies
    print('--Calculating token frequencies')
    sql = """INSERT INTO token_frequency (project_id, token, language, frequency) SELECT
project_id, token, language, count(position) AS frequency FROM tokens GROUP BY project_id,
token, language"""
    cursor.execute(sql)

    #Read tokens
    print('--Removing stopwords')
    query = """SELECT project_id, 1 as ngram_type, token, language, frequency FROM
token_frequency"""
    cursor.execute(query)
    unigram_result = cursor.fetchall()

    #Store results
    store_ngram(unigram_result)

def find_bigrams(text_field,within_distance,min_occurrence):

    ngram_type = 2
    bigram_measures = nltk.collocations.BigramAssocMeasures()
    x = 0
    while x < num_records:
        print(x)

        #Basic data

        this_record = list(text_database[x])
        project = this_record[0]
        rec_language = this_record[2]
        tokens = word_tokenize(this_record[text_field])

        #Find and sort bigrams

        finder = BigramCollocationFinder.from_words(tokens,within_distance)
        finder.apply_freq_filter(min_occurrence)

        if rec_language == 'en':
            finder.apply_word_filter(lambda w: len(w) < 3 or w.lower() in stopwords_en)
        else:
            finder.apply_word_filter(lambda w: len(w) < 3 or w.lower() in stopwords_nl)

        scored = finder.score_ngrams(bigram_measures.raw_freq)
        sorted(bigram for bigram, score in scored)
        bigram_list = list(finder.ngram_fd.items())
        bigram_list.sort(key=lambda item: item[-1], reverse=True)

        #Extract individual items from nested tuples
        bigram_left_item = list(item[0][0] for item in bigram_list)
        bigram_right_item = list(item[0][1] for item in bigram_list)
        bigram_freq = list(item[1] for item in bigram_list)

        #Make list and store bigrams
```

```
        num_bigrams = len(bigram_list)
        bigram_row = []
        y = 0
        while y < num_bigrams:

            list_item1 = str(bigram_left_item[y:y+1])
            list_item2 = str(bigram_right_item[y:y+1])
            list_item3 = str(bigram_freq[y:y+1])
            str_item1 = list_item1[2:len(list_item1)-2]
            if str_item1[0]=='-':
                str_item1 = str_item1[1:len(str_item1)]
            str_item2 = list_item2[2:len(list_item2)-2]
            if str_item2[0]=='-':
                str_item2 = str_item2[1:len(str_item2)]
            str_item3 = list_item3[1:len(list_item3)-1]

            if len(str_item1)<=40 and len(str_item2)<=40:
                bigram_join = ' '.join((str_item1,str_item2))
                row_values = (project,ngram_type,bigram_join,rec_language,int(str_item3))
                bigram_row.append(row_values)

            y += 1

        store_ngram(bigram_row)

        x += 1

def find_trigrams(text_field,within_distance,min_occurrence):

    ngram_type = 3
    trigram_measures = nltk.collocations.TrigramAssocMeasures()
    x = 0
    while x < num_records:
        print(x)

        #Basic data

        this_record = list(text_database[x])
        project = this_record[0]
        rec_language = this_record[2]
        tokens = word_tokenize(this_record[text_field])

        #Find and sort trigrams

        finder = TrigramCollocationFinder.from_words(tokens,within_distance)
        finder.apply_freq_filter(min_occurrence)

        if rec_language == 'en':
            finder.apply_word_filter(lambda w: len(w) < 3 or w.lower() in stopwords_en)
        else:
            finder.apply_word_filter(lambda w: len(w) < 3 or w.lower() in stopwords_nl)
        scored = finder.score_ngrams(trigram_measures.raw_freq)
        sorted(trigram for trigram, score in scored)
        trigram_list = list(finder.ngram_fd.items())
        trigram_list.sort(key=lambda item: item[-1], reverse=True)

        #Extract individual items from nested tuples

        trigram_left_item = list(item[0][0] for item in trigram_list)
        trigram_middle_item = list(item[0][1] for item in trigram_list)
        trigram_right_item = list(item[0][2] for item in trigram_list)
        trigram_freq = list(item[1] for item in trigram_list)

        #Make list and store trigrams

        num_trigrams = len(trigram_list)
        trigram_row = []
        y = 0
```

```
        while y < num_trigrams:

            list_item1 = str(trigram_left_item[y:y+1])
            list_item2 = str(trigram_middle_item[y:y+1])
            list_item3 = str(trigram_right_item[y:y+1])
            list_item4 = str(trigram_freq[y:y+1])
            str_item1 = list_item1[2:len(list_item1)-2]
            if str_item1[0]=='-':
                str_item1 = str_item1[1:len(str_item1)]
            str_item2 = list_item2[2:len(list_item2)-2]
            if str_item2[0]=='-':
                str_item2 = str_item2[1:len(str_item2)]
            str_item3 = list_item3[2:len(list_item3)-2]
            if str_item3[0]=='-':
                str_item3 = str_item3[1:len(str_item3)]
            str_item4 = list_item4[1:len(list_item4)-1]

            if len(str_item1)<=40 and len(str_item2)<=40 and len(str_item3)<=40:
                trigram_join = ' '.join((str_item1,str_item2,str_item3))
                row_values = (project,ngram_type,trigram_join,rec_language,int(str_item4))
                trigram_row.append(row_values)

            y += 1

        store_ngram(trigram_row)

        x += 1

#Main

def main():

    #Global variables

    global conn, cursor, result, num_records, stopwords_en, stopwords_nl, language,
text_database, out_file

    #Connect to sql database

    conn = sqlite3.connect(":memory:")
    cursor = conn.cursor()

    #Write selected fields from a database table into a tab-delimited text file

    out_file = open('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_1_ngrams.csv','w', newline='\n',encoding='utf8')
#  out_file = open('C:/WERK/CBS/RVO/data/1_ngrams.csv','w', newline='\n',encoding='utf8')

    #Read full-text data
    print("Reading full-text data")

#  result = import_text_file('C:/WERK/CBS/RVO/data/cordis_fp7_test.csv','utf8')
    result = import_text_file('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_ket_analyse.csv','utf8')
    num_records = len(result)

    delimiter_fields = '\t'
#  delimiter_fields = '/%/'
    text_database = []
    x=0
    while x < num_records:
        this_record = result[x].split(delimiter_fields)
        project_id = this_record[0]
        text_field = this_record[1]
        language = english_or_dutch(text_field)
        record_values = (project_id, text_field, language)
        text_database.append(record_values)
```

```
        x += 1

    #Import stopword lists
    print("Importing stopword lists")

    stopwords_en = import_text_file('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/python/stopwords_en.txt','utf8')
    stopwords_nl = import_text_file('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/python/stopwords_nl.txt','utf8')
#  stopwords_en = import_text_file('C:/WERK/CBS/RVO/python/stopwords_en.txt','utf8')
#  stopwords_nl = import_text_file('C:/WERK/CBS/RVO/python/stopwords_nl.txt','utf8')

    print('Finding unigrams')
    find_unigrams(1)       #min_occurrence = 1

    print('Finding bigrams')
    find_bigrams(1,3,1)       #text_field: [1]=text_field
                        #within_distance = 3
                        #min_occurrence = 1

    print('Finding trigrams')
    find_trigrams(1,4,1)   #text_field: [1]=text_field
                        #within_distance = 4
                        #min_occurrence = 1

    cursor.close()
    conn.close()

    out_file.close()

    end_time = time.process_time()
    print(end_time)

if __name__ == '__main__':
    main()
```

Ngrams voor een specifieke set (gevalideerde) projecten verzamelen

```
#NGRAM RETRIEVER
#Version 1
#Edwin Horlings
#CBS, February 2019

import sqlite3
from sqlite3 import Error
import time

# Functions

def import_text_file(location,charmap):

    file = open(location, 'r', encoding=charmap)     #Load text file
    rows = file.read().splitlines()                  #read without \n
    file.close()

    return rows

def store_ngram(list_of_values):

    num_records = len(list_of_values)

    selected_text_list = []
    x=0
    while x < num_records:
        this_record = list_of_values[x]
```

```
        text_field = '\t'.join(this_record)
        out_record = (text_field, '\n')
        text_to_write = ''.join(out_record)
        selected_text_list.append(text_to_write)

        x += 1

    selected_text = ''.join(selected_text_list)
    out_file.write(selected_text)

# Main

def find_matched_records():

    print("--reading text files")

    ngrams = import_text_file('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_1_ngrams.csv','utf8')
    selected_recs = import_text_file('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_2_selected_records.csv','utf8')
#  ngrams = import_text_file('C:/WERK/CBS/RVO/data/1_ngrams.csv','utf8')
#  selected_recs = import_text_file('C:/WERK/CBS/RVO/data/2_selected_records.csv','utf8')

    print("--finding matched records")

    num_ngrams = len(ngrams)
    num_selected_recs = len(selected_recs)

    print("--> number of ngrams = ",num_ngrams)
    print("--> number of selected records = ",num_selected_recs)

    x = 0
    while x < num_ngrams:

        this_ngram = ngrams[x].split("\t")
        project_ngram = this_ngram[0]
        ngram = this_ngram[2]

        ngram_row = []
        y = 0
        while y < num_selected_recs:

            this_rec = selected_recs[y].split("\t")
            project_rec = this_rec[0]

            if project_rec == project_ngram:
                ngram_row.append(this_ngram)

            y += 1

        store_ngram(ngram_row)

        x += 1

    out_file.close()

def calculate_statistics():

    print("--preparing database")

    #Connect to sql database

    conn = sqlite3.connect(":memory:")
    cursor = conn.cursor()

    #Create output tables in database
    sql = """DROP TABLE IF EXISTS ngrams"""
    cursor.execute(sql)
```

```
    sql = """CREATE TABLE ngrams (project_id INTEGER, ngram_type INTEGER, ngram TEXT,
language TEXT, frequency INTEGER)"""
    cursor.execute(sql)

    print("--calculating statistics")

#   retrieved_recs = import_text_file('C:/WERK/CBS/RVO/data/3_retrieved.csv','utf8')
    retrieved_recs = import_text_file('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_3_retrieved.csv','utf8')

    num_retrieved = len(retrieved_recs)
    ngram_row = []
    x = 0
    while x < num_retrieved:

        this_ngram = retrieved_recs[x].split("\t")
        ngram_row_add =
(this_ngram[0],this_ngram[1],this_ngram[2],this_ngram[3],this_ngram[4])
        ngram_row.append(ngram_row_add)

        x += 1

    sql = """INSERT INTO ngrams (project_id, ngram_type, ngram, language, frequency)
VALUES (?, ?, ?, ?, ?)"""
    cursor.executemany(sql,ngram_row)

    query = """SELECT language, ngram, ngram_type, count(project_id) AS projects,
sum(frequency) AS total_frequency FROM ngrams GROUP BY language, ngram, ngram_type"""
    cursor.execute(query)
    ngram_result = cursor.fetchall()

    print("--extracting statistics")

    num_stats = len(ngram_result)
    stats_list = []
    x = 0
    while x < num_stats:

        this_record = ngram_result[x]
        stats_row = (this_record[0], this_record[1], str(this_record[2]), str(this_record[3]),
str(this_record[4]))
        stats_list.append(stats_row)

        x += 1

    print("--storing results")

    store_ngram(stats_list)

    out_file.close()

    cursor.close()
    conn.close()


def main():

    #Global variables

    global ngrams, selected_recs, out_file

    print("Find records in ngrams that match record ids in selected_recs")

    out_file = open('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_3_retrieved.csv','w', newline='\n',encoding='utf8')
#   out_file = open('C:/WERK/CBS/RVO/data/3_retrieved.csv','w', newline='\n',encoding='utf8')
    find_matched_records()
```

```
    print("Calculate statistics for all ngrams with frequency > 1")

    out_file = open('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_4_stats_retrieved.csv','w',
newline='\n',encoding='utf8')
#  out_file = open('C:/WERK/CBS/RVO/data/4_stats_retrieved.csv','w',
newline='\n',encoding='utf8')
    calculate_statistics()

    end_time = time.process_time()
    print(end_time)


if __name__ == '__main__':
    main()
```

Zoeken naar WBSO-projecten waarin de ngrams uit de set (gevalideerde) projecten voorkomen

```
#NGRAM SEARCH RECORDS
#Version 1
#Edwin Horlings
#CBS, February 2019

import sqlite3
from sqlite3 import Error
import time

# Functions

def import_text_file(location,charmap):

    file = open(location, 'r', encoding=charmap)     #Load text file
    rows = file.read().splitlines()                  #read without \n
    file.close()

    return rows

def store_ngram(list_of_values):

    num_records = len(list_of_values)

    selected_text_list = []
    x=0
    while x < num_records:
        this_record = list_of_values[x]
        out_record = (this_record, '\n')
        text_to_write = ''.join(out_record)
        selected_text_list.append(text_to_write)

        x += 1

    selected_text = ''.join(selected_text_list)
    out_file.write(selected_text)

def main():

    #Global variables

    global ngrams, selected_recs, out_file

    print("Reading text files")
```

```
    ngrams = import_text_file('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_1_ngrams.csv','utf8')
    search_terms = import_text_file('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_5_search_terms.csv','utf8')

#   ngrams = import_text_file('C:/WERK/CBS/RVO/data/1_ngrams.csv','utf8')
#   search_terms = import_text_file('C:/WERK/CBS/RVO/data/5_search_terms.csv','utf8')

    out_file = open('E:/Pilotproject_Textmining_KETs_WBSO/WBSO-
projecten/text_mining_KETs/data/wbso_6_matched_projects.csv','w',
newline='\n',encoding='utf8')
#   out_file = open('C:/WERK/CBS/RVO/data/6_matched_projects.csv','w',
newline='\n',encoding='utf8')

    num_ngrams = len(ngrams)
    num_search_terms = len(search_terms)

    print("--> number of ngrams = ",num_ngrams)
    print("--> number of search terms = ",num_search_terms)

    print("Looking for matches")

    matching_projects = []
    x = 0
    while x < num_ngrams:

        this_ngram = ngrams[x].split("\t")
        project_ngram = str(this_ngram[0])
        ngram = this_ngram[2]

        y = 0
        while y < num_search_terms:

            search_term_to_match = search_terms[y]

            if search_term_to_match == ngram:
                if project_ngram not in matching_projects:
                    matching_projects.append(project_ngram)

            y += 1

        x += 1

    print("Storing results")

    store_ngram(matching_projects)

    out_file.close()

    end_time = time.process_time()
    print(end_time)


if __name__ == '__main__':
    main()
```

Dit is een publicatie van:

Rijksdienst voor Ondernemend Nederland
Hanzelaan 310 | 8017 JK Zwolle
Postbus 10073 | 8000 GB Zwolle
T +31 (0) 88 042 42 42
E klantcontact@rvo.nl

Publicatienummer: RVO-108-1901-RP/CORP

De Rijksdienst voor Ondernemend Nederland (RVO.nl) is een agentschap van het Ministerie van Economische Zaken en Klimaat. RVO.nl voert beleid uit voor verschillende ministeries en decentrale overheden als het gaat om duurzaam, agrarisch, internationaal en innovatief ondernemen. RVO.nl is hèt aanspreekpunt voor bedrijven, kennisinstellingen en overheden als het gaat om informatie en advies, financiering, netwerken en wet- en regelgeving.