



Documentatie typeerproces beroep

Sue Westerman, 03-04-2018

Inhoudsopgave

1. Inleiding	4
2. Typeerproces, twee varianten	5
3. Procesverloop variant 1	7
3.1 Typeerstappen	7
3.2 Input	8
3.3 Output	10
3.4 Kwaliteit	11
4. Procesverloop variant 2	12
4.1 Typeerstappen	12
4.2 Input	13
4.3 Output	13
4.4 Kwaliteit na inzetten handmatig typen	14
4.5 Kwaliteit na volledige automatisch typering	16
Bijlage 1. Meten kwaliteit volledig automatisch typen	18
Bijlage 2. Kwaliteit typeerproces volledig automatisch BRC 2014	21
Bijlage 3. Kwaliteit typeerproces volledig automatisch ISCO 2008	29
Bijlage 4. Kruistabellen 1^e aggregatieniveau BRC 2014 en ISCO 2008	39
Bijlage 5. Schematisch overzicht en praktijkvoorbeelden	41

1. Inleiding

Het CBS beschikt over de mogelijkheid om antwoorden in persoonsenquêtes op vragen over beroep en de belangrijkste werkzaamheden te typeren naar beroepen volgens de internationale beroepenclassificatie ISCO 2008 (International Standard Classification of Occupations 2008). Tijdens het typeren wordt de enorme verscheidenheid aan mogelijke beroepsomschrijvingen gestructureerd en voorzien van codes van standaard beroepenindelingen. De in het veld verzamelde informatie over beroep is daarna geschikt voor analyses.

In deze nota wordt een toelichting gegeven op de manier waarop het typeerproces van beroep is ingericht¹. Beroep wordt getypeerd via twee onderling verschillende processen. Het verschil ligt in de mate van detail en kwaliteit die wordt opgeleverd en daarmee samenhangend het aandeel dat automatisch of handmatig wordt getypeerd. Eén procesvariant is gericht op het opleveren van de ISCO 2008 op 4 digits met voldoende kwaliteit in een vaste verhouding tussen automatisch en handmatig typeren, in de andere procesvariant wordt afhankelijk van gewenste kwaliteit en detail meer of minder handmatig getypeerd. In het eerste hoofdstuk worden deze twee procesvarianten verder toegelicht, en wordt ingegaan op verschillen in de uitgangspunten van de beide processen. In de twee volgende hoofdstukken wordt per procesvariant uitgebreid ingegaan op de typeerstappen per variant, de input die wordt gebruikt, de output die wordt opgeleverd en de kwaliteit daarvan. Bijlage 1 gaat in op de meting van kwaliteit wanneer beroep volledig automatisch wordt getypeerd, bijlage 2 en 3 laten de kwaliteit en verdeling zien na volledig automatische typering over de categorieën van de beroepenindelingen Beroepenindeling ROA-CBS 2014 (BRC 2014)² en de International Standard Classification of Occupations 2008 (ISCO 2008)³. In bijlage 4 wordt door middel van een kruistabel inzicht gegeven in de verdeling van de typering over het hoogste aggregatieniveau van de beide indelingen volgens de beide typeerprocessen. In bijlage 5 wordt het verloop van de beide procesvarianten geïllustreerd aan de hand van een schematische weergave en een aantal praktijkvoorbeelden.

¹ Met dank aan Elena Grigorieva, Birgit van Gils, Martijn Souren en Hendrika Lautenbach die door het geven van commentaar en stellen van vragen hebben bijgedragen aan de totstandkoming van de documentatie

² <https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/onderwijs-en-beroepen/beroepenclassificatie--isco-en-sbc-->

³ <http://www.ilo.org/public/english/bureau/stat/isco/isco08/>

2. Typeerproces, twee varianten

Voor het typeren van informatie over beroep verzameld in persoonsenquêtes zijn twee processen ontworpen die ieder een eigen verloop hebben. In beide processen wordt in de eerste twee stappen automatisch getypeerd op basis van beroep en de combinatie beroep en werkzaamheden. De mate waarin aanvullende hulpinformatie over beroep wordt gebruikt tijdens automatisch typeren, en de wijze waarop handmatig typeren wordt toegepast verschilt tussen de beide processen.

De beide varianten zijn niet gelijktijdig ontwikkeld, en beantwoorden aan verschillende behoeften met betrekking tot het meten van de variabele beroep. Het eerst ontwikkelde typeerproces is er op gericht de International Standard Classification of Occupations 2008 (ISCO 2008) op het meest gedetailleerde niveau op te leveren met voldoende kwaliteit. Dit proces verloopt deels automatisch en deels handmatig. Het tweede typeerproces is meer flexibel. Er kan in deze variant volledig automatisch worden getypeerd en het inzetten van handmatig typeren is optioneel. Dit proces is er op gericht om voldoende kwaliteit te leveren van het voor publicatie- en analysedoeleinden gewenste aggregatieniveau, maar niet meer dan dat.

De eerste variant van het typeerproces is ontwikkeld tijdens het herontwerp van de persoonsenquêtes dat in 2011 en 2012 heeft plaatsgevonden. Algemeen doel van het herontwerp was om persoonsenquêtes meer kosteneffectief te organiseren. Voor het typeerproces van beroep hield het in dat het aan de volgende eisen moest voldoen:

1. Het typeerproces moest geschikt zijn voor waarneming van data over beroep verzameld via drie verschillende interviewmodes: CAPI, CATI en CAWI, op dezelfde manier in de 3 verschillende modes en in verschillende onderzoeken (EBB, NEA, ZEA, AKO etc.). CAWI werd tijdens het herontwerp voor het eerst geïntroduceerd in de waarneming van beroep via persoonsenquêtes.
2. Het typeerproces moest alleen nog van generieke softwaretools gebruik maken om te kunnen besparen op ontwikkelkosten en kosten van onderhoud en beheer van maatwerktools en bij voorkeur geschikt zijn voor zowel handmatig als automatisch typeren van beroep. Hiervoor werd de tool Cascot gekozen ontwikkeld door de universiteit van Warwick⁴.
3. Het typeerproces moest minimaal de ISCO 2008 op 3 digits met voldoende kwaliteit opleveren, omdat het CBS hiertoe wettelijk⁵ verplicht is. Er werd besloten om de ISCO op het laagste 4^e aggregatieniveau op 4 digits op te leveren om ten behoeve van andere onderzoeken te compenseren voor het verlies aan detailinformatie over beroep. De Standaard Beroepen Classificatie 1992 waarin 1211 beroepen worden onderscheiden zou namelijk niet meer gebruikt kunnen worden na herontwerp.
4. Het typeerrendement van automatische typering op 4 digits met voldoende kwaliteit moest bij voorkeur hoger, maar minimaal gelijk zijn aan wat typeerproces⁶ dat tot dan toe opleverde, dit was minimaal 60%.

De tweede variant van het typeerproces is in 2016 ontwikkeld ten behoeve van een maatwerkopdracht⁷ voor het onderzoeksbureau Lifelines. Het werd zo ingericht dat het proces breder toepasbaar zou kunnen zijn. Er werden aan dit proces andere eisen gesteld:

⁴ <https://warwick.ac.uk/fac/soc/ier/software/cascot/>

⁵ Het betreft de wettelijk verplichting tot leveren van data verzameld in de Enquete Beroepsbevolking op kwartaal- en jaarbasis aan Eurostat.

⁶ Zie Computer assisted coding by interviewers, Wim Hacking, John Michiels, Saskia Janssen-Jansen, CBS 2008

⁷ Lifelines heeft het CBS in 2015 gevraagd om een proces te ontwikkelen waarmee de verzamelde informatie over beroep (en werkzaamheden) volledig automatisch kan worden getypeerd volgens de International Standard Classification of Occupations 2008 (ISCO 2008). In 2016 is dit typeerproces opgeleverd en via dit proces getypeerde data geleverd conform het maatwerkverzoek.

1. Het moest mogelijk zijn antwoorden op vragen naar beroep en/of werkzaamheden volledig automatisch te voorzien van geldige ISCO-codes, ook indien niet waargenomen via de standaard vraagstelling van beroep van het CBS.
2. Opgeleverde data moest per categorie van het 1e tot en met 3e aggregatieniveau van de ISCO 2008 en BRC 2014 zijn voorzien van een indicatie van de kwaliteit.
3. Handmatig typen zou flexibel ingezet moeten kunnen worden, namelijk alleen indien daaraan behoefte is om de kwaliteit te verbeteren van het 1e, 2e of 3e aggregatieniveau.

Dit proces is met ingang van 2018 door het CBS in gebruik genomen.

De keuze voor de ene dan wel andere variant hangt af van de outputwens van de opdrachtgever. Wanneer een opdrachtgever getypeerde data over beroep wil gebruiken op het meest gedetailleerde 4e aggregatieniveau van de ISCO 2008 met voldoende kwaliteit of wanneer beroep een doelvariabele is, wordt geadviseerd om gebruik te maken van de meest geavanceerde eerst beschreven variant van het typeerproces. Het is hiervoor nodig om de standaard beroepenvraagstelling van het CBS te gebruiken, zie paragraaf 3.2. Dan kan optimaal gebruik worden gemaakt van aanvullende variabelen en handmatige typeercapaciteit worden ingezet ten behoeve van een acceptabel kwaliteitsniveau van typering op 4 digits. Indien voor de analysedoeleinden minder detail nodig is of wanneer beroep geen doel- maar een achtergrondvariabele is, kan worden volstaan met de tweede procesvariant om daarmee op kosten te besparen. In theorie is het mogelijk om hiervoor een eigen beroepenvraagstelling te gebruiken, zie paragraaf 4.2.

In tabel 1 staan de aggregatieniveaus, het aantal digits van de codering per aggregatieniveau en aantallen categorieën daarbinnen van de ISCO 2008 en BRC 2014 naast elkaar gepresenteerd. Beide indelingen hebben een hiërarchische structuur. In de ISCO 2008 hebben de codes op het hoogste aggregatieniveau (major groups) 1 digit, ieder lager aggregatieniveau heeft steeds 1 digit meer. De BRC 2014 heeft op het hoogste niveau (beroepsklassen) 2 digits, en voor het 2^e aggregatieniveau 1 digit en het 3^e aggregatieniveau 2 digits meer. De BRC 2014 beroepen zijn gelijk aan de unit groups van de ISCO 2008 en hebben 4 digits. In paragraaf 3.3 en 4.3 wordt voor de beide procesvarianten toegelicht hoe de output beschikbaar wordt gesteld.

Tabel 1 De aggregatieniveaus van de ISCO 2008 en de BRC 2014, aantal groepen per aggregatieniveau en codering van aantal digits per aggregatieniveau. Per aggregatieniveau de aanbevolen te gebruiken procesvariant.

Aanbevolen procesvariant	ISCO 2008			BRC 2014			
	Aggregatie niveau	Indeling	Codering Aantal digits	Aantal groepen	Indeling	Aantal groepen	Codering Aantal digits
Variant 2	1	Major groups	1 dig	10	Beroepsklassen	13	2 dig
Variant 2	2	Sub-major groups	2 dig	43	Beroepssegmenten	41	3 dig
Variant 2	3	Minor groups	3 dig	130	Beroepsgroepen	114	4 dig
Variant 1	4	Unit groups	4 dig	436	Beroepen	436	4 dig

In beide typeerprocessen wordt Cascot gebruikt voor automatisch en handmatig typen van beroep. In het programma wordt een classificatiefile gebruikt waarin een index of zoeklijst van beroepenbenamingen zijn opgenomen en er zijn regels opgesteld ter bevordering van de kwaliteit van de typering, er kunnen bijvoorbeeld afkortingen of synoniemen worden opgegeven. Tijdens het typen van beroep worden alle typering door het programma Cascot voorzien van een score. De score is een maat voor de waarschijnlijkheid dat de toegekende code correct is. Hoe hoger de score hoe beter de kwaliteit. De score is in het typeerproces een belangrijk sturingsvariabele die in de beide processen verschillend wordt toegepast, zie paragraaf 3.1 en 4.1. In paragraaf 3.4 en 4.4 wordt ingegaan op de kwaliteit van de uiteindelijk opgeleverde output.

3. Procesverloop variant 1

3.1 Typeersteps

Procesvariant 1 is opgedeeld in 4 stappen. In de eerste twee stappen tijdens automatisch typen en in de 4^e stap tijdens handmatig typen wordt het programma Cascot gebruikt. In Cascot wordt een classificatiefile gebruikt waarin een index of zoeklijst met beroepenbenamingen is opgenomen en ook een set van zoekregels die helpen de juiste code toe te kennen aan de getypte tekst. Er wordt in deze procesvariant tijdens handmatig typen een andere classificatiefile gebruikt dan in de eerste twee stappen tijdens automatisch typen.

3.1.1 Stap 1: inputvariabele beroep

In de eerste stap van het typeerproces wordt automatisch getypeerd op grond van de antwoorden op de vraag naar beroep. De lijst van voorbeeldberoepen per ISCO-unit group die te raadplegen is in de handleiding⁸ van de ISCO 2008 van de International Labour Organisation (ILO) is uitgangspunt in de afbakening van beroepsaanduidingen die zonder werkzaamheden of andere hulpinformatie erbij te betrekken getypeerd kunnen worden. Het betreffen beroepsaanduidingen waarvan de taken die kenmerkend zijn voor een bepaalde ISCO-unit group vrijwel altijd deel uitmaken van het takenpakket. Zo kan bijvoorbeeld de beroepsomschrijving 'bioloog' ook zonder de werkzaamheden daarbij te betrekken van een ISCO-code worden voorzien, maar de omschrijving 'wetenschapper' niet.

In deze stap worden records voorzien van een geldige unit group code van de ISCO 2008. Dit is het laagste aggregatieniveau van de ISCO 2008 en de code heeft 4 digits. Er wordt in deze stap niet getypeerd met een vangnetcode⁹. Bij bepaalde beroepen wordt er een afleidingscode¹⁰ toegekend.

De records die zijn getypeerd met een score van 40 of meer en niet met een afleidingscode komen in de output terecht, de rest gaat door naar de vervolgstap. Rond de 60% van alle te typen antwoorden wordt in deze stap getypeerd, zie tabel 2.

3.1.2 Stap 2: inputvariabelen beroep en werkzaamheden

In de tweede stap worden antwoorden op de vragen naar beroep en belangrijkste werkzaamheden gecombineerd tot een tekstveld dat de input vormt voor de automatische typeringen in deze stap. Zo kan bijvoorbeeld het antwoord 'monteur' dat te globaal is om in de 1^e typeerstep te typen na combinatie met een omschrijving van werkzaamheden als 'auto's repareren' wel getypeerd worden.

De records die zijn getypeerd met geldige ISCO-code met een score van 70 of meer komen in de output terecht. Rond de 2% van alle te typen antwoorden wordt in deze stap getypeerd, zie tabel 2.

Het deel dat een afleidingscode toegekend heeft gekregen gaat door naar stap 3. Als in beide stappen een afleidingscode heeft gekregen krijgt de afleidingscode van de 1^e stap prioriteit.

3.1.3 Stap 3: afleiding met behulp van leidinggevende taken en SBI

In de derde stap wordt een afleidingsschema doorlopen waarmee de beroepenantwoorden die in de eerste twee stappen zijn voorzien van een afleidingscode naar ISCO-codes kunnen worden afgeleid onder bepaalde voorwaarden van hulpvariabelen. In principe wordt dit toegepast bij beroepenantwoorden die te summier zijn

⁸ <http://www.ilo.org/public/english/bureau/stat/isco/isco08/>

⁹ Een vangnetcode is een code op 4 digits waarbij de code van het hogere aggregatieniveau van de ISCO 2008 is aangevuld met nullen. In de omschrijving is het beroep aangevuld met de term zonder nadere differentiatie (z.n.d.). Bijvoorbeeld het antwoord militair wordt getypeerd als 0000, en heeft als omschrijving Militaire beroepen z.n.d.

¹⁰ Een afleidingscode is een hulpcode met 5 digits die via de defaultcoding rules van Cascot aan een specifieke selectie beroepen toegekend wordt. De afleidingscodes zijn om die reden ook in de zoekindex en in de structuur van de classificatiefile van Cascot opgenomen. De afleidingscodes vormen input voor typeerstep 3.

omschreven om op grond daarvan een ISCO-code te kunnen toekennen, bijvoorbeeld directeur, ondernemer, accountmanager, sales manager of monteur. De SBI en antwoorden op de vragen naar leidinggevende taken worden dan ingezet om op basis hiervan een ISCO-code te kunnen toekennen.

In deze stap worden ook records die in stap 1 of in stap 2 zijn getypeerd met een ISCO-code van major group 1 'Managers' doorgestuurd naar handmatig typen als bekend is dat de persoon geen leiding geeft.

Records waar de hulpvariabelen voldoen aan de gespecificeerde voorwaarden worden getypeerd met een 4-digit ISCO code en komen in de output, de rest gaat door naar de vervolgstap, handmatig typen.

Rond de 12% van alle te typen antwoorden wordt in deze stap getypeerd, zie tabel 2.

3.1.4 Stap 4: handmatig typen

In de vierde stap worden alle records handmatig getypeerd die niet in de voorafgaande stappen van een code kunnen worden voorzien omdat in stap 1 de score lager dan 40 is, in stap 2 lager dan 70 of de afleidingsregels in stap 3 niet van toepassing zijn.

Typeurs gebruiken voor de typeringen in aanvulling op de al gebruikte variabelen nog informatie over het aantal personen waaraan wordt leidinggegeven en het hoogst gevolgde onderwijsniveau. Tijdens het handmatig typen is het toegestaan om op een hoger aggregatieniveau te typen van de ISCO-classificatie. Er worden hiervoor vangnetcodes gebruikt.

Rond de 28% wordt in deze typeerstep handmatig getypeerd.

Tabel 2. Verdeling over de typeerstappen, EBB 201612 tm EBB 201707, ongewogen, 40 duizend records.

Typeerstep	Aandeel t.o.v. totaal
	%
Automatisch	
1-beroep	58
2-beroep & werkzaamheden	2
3-hulpvariabelen	12
Totaal automatisch	72
Handmatig	
4-handmatig typen	28
Totaal	100

Er wordt in deze stap een andere classificatiefile gebruikt dan in de eerste 2 stappen van het typeerproces.

3.2 Input

Procesvariant 1 kan alleen worden gebruikt indien beroep is waargenomen via de standaard CBS-vraagstelling. De volgende variabelen uit deze vraagstelling¹¹ worden gebruikt:

- Tijdens automatisch typen worden gebruikt: Beroep, VoornWzh, Leiding, UitsLeid, DeelLeid, PersBeleid, StratBeleid en de SBI-code.
- Tijdens handmatig typen worden in aanvulling daarop nog gebruikt: Afl_HgstNivGev, N_LeidW, BedrOmvZ.

¹¹ De vraagstelling is beschikbaar op de website van het CBS onder methoden/onderzoek/aanvullende onderzoeksbeschrijvingen.

De CBS-beroepenvraagstelling is erop ingericht om beroep volgens de ISCO 2008 te kunnen waarnemen. De vraagstelling volgt aanbevelingen¹² van de ILO met betrekking tot waarneming van beroep in enquêtes en bevat open vragen over het beroep en de belangrijkste werkzaamheden. Daarnaast stelt het CBS in de beroepenvraagstelling aan personen die leiding geven aan personeel nog een aantal extra vragen over het soort leidinggevende taken ter verbetering van de typering van de leidinggevendenden.

Eerst wordt de vraag gesteld naar het beroep van de respondent.

Beroep

Welk beroep of welke functie oefent (\$A: u \$B: hij \$C: zij) uit?

>>ENQ.: Probeer in de omschrijving zo specifiek mogelijk te zijn, bijvoorbeeld door een specialisme of niveau op te geven.

<i>Dus niet:</i>	<i>Maar liever:</i>
<i>Manager</i>	<i>Manager automatisering, Manager zorg, Financieel Manager</i>
<i>Verpleegkundige</i>	<i>Psychiatrisch verpleegkundige, Verpleegkundige niveau 4, Verpleegkundige op de spoedeisende hulp</i>
<i>Monteur</i>	<i>Automonteur, Monteur elektrotechniek, Machinemonteur<<</i>

Daarna volgen een aantal gesloten vragen naar de leidinggevende taken.

Leiding

Geeft (\$A: u \$B: hij \$C: zij) in (\$A: uw \$B: zijn \$C: haar) beroep of functie leiding aan medewerkers of personeel?
[TJaNee]

N_LeidW

Aan hoeveel personen?

- | | |
|----------------|-------------------|
| 1. 1 – 4 | <i>[N_1tm4]</i> |
| 2. 5 – 9 | <i>[N_5tm9]</i> |
| 3. 10 – 19 | <i>[N_10tm19]</i> |
| 4. 20 – 49 | <i>[N_20tm49]</i> |
| 5. 50 – 99 | <i>[N_50tm99]</i> |
| 6. 100 of meer | <i>[N100meer]</i> |

Aan zelfstandigen wordt in blok bedrijf gevraagd of ze personeel in dienst hebben en zo ja hoeveel personeelsleden ze in dienst hebben. Vervolgens doorlopen ze dezelfde vragen als de werknemers.

Uitsleid

Geeft (\$A: u \$B: hij \$C: zij) uitsluitend leiding of verricht (\$A: u \$B: hij \$C: zij) daarnaast ook dezelfde werkzaamheden als het personeel of de medewerkers waaraan (\$A: u \$B: hij \$C: zij) leiding geeft?

- | | |
|---|-------------------|
| 1. Geeft uitsluitend leiding | <i>[Uitsluit]</i> |
| 2. Naast leiding geven dezelfde werkzaamheden als personeel/medewerkers | <i>[ZelfWrkz]</i> |

DeelLeid

Waaruit bestaat het grootste deel van (\$A: uw \$B: zijn \$C: haar) werkzaamheden?

- | | |
|-------------------------|-------------------|
| 1. Leiding geven | <i>[Leiding]</i> |
| 2. Andere werkzaamheden | <i>[AndrWrkz]</i> |

Alleen als het grootste deel van de werkzaamheden leiding geven is wordt er daarna gevraagd naar de beslissingsbevoegdheden.

PersBeleid

Heeft (\$A: u \$B: hij \$C: zij) de bevoegdheid om beslissingen te nemen over personele zaken zoals het aannemen van personeel of het geven van een loonsverhoging?
[TJaNee]

StratBeleid

Heeft (\$A: u \$B: hij \$C: zij) de bevoegdheid om beslissingen te nemen met betrekking tot het financieel of strategisch beleid van de

¹² <http://www.ilo.org/public/english/bureau/stat/isco/isco08/>

organisatie, zoals het budget of het meerjarenplan?
[TJaNee]

Als laatste krijgen respondenten een vraag naar belangrijkste werkzaamheden.

VoornWzh

Wat zijn (§1: naast leidinggeven) de belangrijkste werkzaamheden die (§A: u §B: hij §C: zij) verricht?

>>ENQ.: Probeer in de omschrijving zo specifiek mogelijk te zijn.

<u>Dus niet:</u>	<u>Maar liever:</u>
Adviseren	Particulieren adviseren over hypotheek, Scholieren adviseren bij vervolgstudie, Bedrijven juridisch advies geven
Administratie	Boekhouden, Leerlingenadministratie bijhouden, factureren, gegevens invoeren
Verzorgen	Verzorgen van kinderen, Verzorgen van ouderen in thuiszorg, Verzorgen van gehandicapten<<

De open vragen naar beroep en werkzaamheden zijn bewust zo ontworpen dat respondenten ertoe worden aangezet zo specifiek mogelijk hun beroep en werkzaamheden te omschrijven om te voorkomen dat beroepen door een gebrek aan informatie niet getypeerd kunnen worden.

Naast de variabelen die in het blok beroep worden waargenomen wordt er in het typeerproces ook gebruik gemaakt van informatie over de bedrijfsactiviteit (SBI), bedrijfsomvang en niveau van het hoogst gevolgde onderwijs. Bedrijfsactiviteit wordt gebruikt tijdens automatisch typeren in stap 3 van het typeerproces, de bedrijfsomvang en niveau van het hoogst gevolgde onderwijs wordt alleen gebruikt in de 4^e typeerstep tijdens handmatig typeren.

Variabele onderwijsniveau:

Afl_HgstNivGev

*Afleiding hoogst gevolgde opleiding

1. Lbo, vso (lts, leao, vbo, huishoudschool, ambachtsschool)	[LBO]
2. Vmbo, lwoo (inclusief theoretische leerweg)	[VMBO]
3. Mavo (ulo, mulo)	[Mavo]
4. Havo (mms)	[Havo]
5. Vwo, gymnasium, atheneum (hbs, lyceum)	[VWO]
6. Mbo (mts, meao, middenstandsdiploma, pdb, mba)	[MBO]
7. Hbo (hts, heao, kweekschool, associate degree)	[HBO]
8. Universitaire opleiding, inclusief postdoctorale opleidingen en promotieonderzoek,	[Univ]
9. Een andere (bedrijfs)opleiding of cursus	[Cursus]

Variabele bedrijfsomvang in blok bedrijf:

BedrOmvZ[i]

(§1: Hoeveel mensen (§1: §A: heeft u §B: heeft hij §C: heeft zij) dan ongeveer in dienst?

§2: (§A: Uzelf meegeteld §B: Als u hem meetelt §C: Als u haar meetelt), hoeveel mensen (§A: heeft uw partner §B: heeft zijn partner §C: heeft haar partner) dan ongeveer in dienst?

§3: (§A: Uzelf meegeteld §B: Als u hem meetelt §C: Als u haar meetelt), hoeveel mensen (§A: hebben uw (schoon)ouders §B: hebben zijn (schoon)ouders §C: hebben haar (schoon)ouders) dan ongeveer in dienst?)

1. 1	[N_1]
2. 2 – 4	[N_2tm4]
3. 5 – 9	[N_5tm9]
4. 10 – 19	[N_10tm19]
5. 20 – 49	[N_20tm49]
6. 50 – 99	[N_50tm99]
7. 100 of meer	[N100meer]

Variabele bedrijfsomvang wordt iets gedetailleerder uitgevraagd dan de antwoordcategorieën van N_LeidW daarom worden voor handmatig typeren de categorieën 1. 1 en 2. 2-4 samengevoegd tot een categorie leidinggeven aan 1-4 personen.

3.3 Output

In procesvariant 1 worden standaard ISCO-2008 codes opgeleverd van het laagste aggregatieniveau van de classificatie, de ISCO unit groups. Er wordt in dit typeerproces alleen tijdens handmatig typeren getypeerd op een hoger aggregatieniveau van de ISCO 2008 in het geval de te typeren input te weinig informatie bevat om in meer detail te typeren. De codes die worden opgeleverd zijn te schakelen naar de major-, sub-major en minor groups van de ISCO 2008, of de beroepsklassen, -segmenten, -groepen van de BRC 2014¹³.

Uit tabel 3 blijkt dat na doorlopen van de typeerstappen 96,3% is getypeerd met een ISCO-code op 4 digits, en slechts 3,7% op een hoger aggregatieniveau van de ISCO 2008 met een vangnetcode is getypeerd. Het aandeel dat wordt getypeerd met beroep onbekend is laag, maar 1,5%.

Tabel 3. Aandeel getypeerd met ISCO 2008 codes op hoger aggregatieniveau met een vangnetcode of met unit group code of met beroep onbekend na doorlopen van de typeerstappen van procesvariant 1, EBB 2016 voor weging, 92 duizend records

Aggregatieniveau ISCO 2008	Aandeel t.o.v. totaal %
Vangnet major groups	1,5
Vangnet sub-major groups	1,3
Vangnet minor groups	0,9
Unit groups	96,3
Onbekend	1,5
Totaal	100

3.4 Kwaliteit

De kwaliteit van procesvariant 1 wordt gemeten door jaarlijks steekproefsgewijs voor een paar duizend records handmatig te controleren of de toegekende 4-digit codes correct zijn gegeven de definitie van de unit groups volgens de handleiding van de ISCO 2008 en beschikbare informatie die tijdens typeren kan worden gebruikt.

Ten opzichte van het totaal van alle typering mag het aandeel met een onjuiste 4-digit code maximaal 5% zijn, per outputcategorie waarover in reguliere statistieken op Statline wordt gepubliceerd (3e aggregatieniveau van de ISCO of BRC) mag het aandeel onjuist maximaal 10% zijn. De gevonden fouten kunnen leiden tot aanpassing in het automatisch typeerproces.

Resultaten worden jaarlijks besproken met diverse betrokkenen van het typeerproces: programmeurs, procescoördinatoren, projectleiders van de verschillende onderzoeken en de beroepentypeurs. Uitwisseling van de ervaringen kan nog leiden tot extra aanvullingen in de index of regels van Cascot of wijzigingen in het typeerproces.

¹³ Beschikbaar op de website van het CBS onder methoden/classificaties/beroepenclassificaties/codelijsten en beroepenindex.xls of in SSB-omgeving via ISCOVREFVV.txt

4. Procesverloop variant 2

4.1 Typeerstappen

Procesvariant 2 is opgedeeld in 3 stappen, waarbij de 3^e typeerstep 'handmatig typen' optioneel is. Er wordt in dit proces getypeerd in Cascot met een classificatiefile. Deze classificatiefile komt overeen de classificatiefile die in procesvariant 1 wordt gebruikt tijdens handmatig typen.

4.1.1 Stap 1: inputvariabele beroep

In de eerste stap van het typeerproces worden de antwoorden op de vraag naar beroep automatisch getypeerd. In deze stap worden alle records voorzien van een geldige unit group code van de ISCO 2008.

Alle records gaan door naar de vervolgstap.

4.1.2 Stap 2: inputvariabelen beroep en werkzaamheden

In de tweede stap worden antwoorden op de vragen naar beroep en belangrijkste werkzaamheden gecombineerd tot een tekstveld dat de input vormt voor de automatische typering in deze stap. Ook in deze stap worden alle records voorzien van een geldige unit group code van de ISCO 2008.

Na de tweede typeerstep wordt de score van de records getypeerd in stap 1 en stap 2 onderling vergeleken. De typering met de hoogste score komt in de output terecht. Als de scores gelijk zijn wordt de code van stap 1 overgenomen. Records die in de 1^e stap met beroep onbekend zijn getypeerd en in de 2^e stap met een geldige ISCO-code komen alleen in de output als de score in de tweede stap 40 of hoger is. De rest krijgt de ISCO-code voor beroep onbekend.

De in deze eerste twee stappen getypeerde dataset wordt in de output database opgenomen als er geen behoefte is om de kwaliteit te verbeteren en hiervoor handmatig typen in te zetten.

Tabel 4 laat zien dat indien volledig automatisch wordt getypeerd 83% wordt getypeerd op basis van de antwoorden op de vragen naar het beroep en 17% op basis van de combinatie van de antwoorden op de vragen naar beroep en werkzaamheden. De verdeling over typering op basis van beroep alleen, of op basis van de combinatie van beroep en werkzaamheden wijkt aanzienlijk af van procesvariant 1 (tabel 2) omdat in die variant alleen automatische getypeerde records op 4 digits met voldoende kwaliteit in de output terecht komen. De afbakening van het deel van de in procesvariant 2 automatisch getypeerde records dat voldoende kwaliteit heeft wordt na doorlopen van de 1^e 2 typeerstappen gedaan, zie verder paragraaf 4.1.3.

Tabel 4 verdeling over typeerstappen ten opzichte van het totaal na volledig automatisch typen, EBB 2016 ongewogen, 92 duizend records

Typeerstep	Aandeel t.o.v. totaal
	%
Automatisch	
1-beroep	83
2-beroep & werkzaamheden	17
Totaal	100

4.1.3 Stap 3: optioneel handmatig typen

Indien de opdrachtgever heeft aangegeven handmatig typen te willen gebruiken om voldoende kwaliteit van alle categorieën binnen het gewenste aggregatieniveau van de beroepenindeling te behalen wordt een selectie gemaakt van de categorieën waarvan de kwaliteit na volledig automatische typering onvoldoende is (zie paragraaf 4.4). Van die categorieën wordt vervolgens het deel handmatig getypeerd dat in de eerste twee

stappen is getypeerd met een score lager dan 40¹⁴. In het maken van deze selectie wordt rekening gehouden met de behoefte aan detail bij de opdrachtgever zodat er niet meer handmatig wordt getypeerd dan nodig.

Typeurs hebben tijdens handmatig typeren de beschikking over de variabelen die ook gebruikt worden tijdens handmatig typeren in procesvariant 1, dus SBI, leidinggevende taken, aantal ondergeschikten en onderwijsniveau.

De records die na doorlopen van de eerste twee typeerstappen zijn getypeerd met beroep onbekend worden in procesvariant 2 niet meer handmatig getypeerd. De reden hiervoor is dat het aandeel met beroep onbekend slechts 4% is ten opzichte van het totaal (zie tabel 5). In vergelijking met het procesvariant 1 is dit aandeel wel hoger (zie tabel 3), maar de extra records die in variant 2 met onbekend zijn getypeerd komen indien getypeerd volgens procesvariant 1 in alle beroepsklassen of major groups voor, en zijn niet geconcentreerd in slechts enkele daarvan (zie bijlage 4). Dit geldt ook voor de verdelingen van de meer gedetailleerde aggregatieniveaus van de ISCO 2008 of BRC 2014. Door de onbekenden niet handmatig te typeren is het aandeel met onbekend in procesvariant 2 wel hoger, maar zal niet ten koste gaan van de kwaliteit van het overige deel met bekende ISCO-code.

4.2 Input

In procesvariant 2 wordt tijdens automatisch typeren alleen gebruik gemaakt van variabelen beroep en werkzaamheden, andere hulpinformatie wordt ingezet tijdens handmatig typeren indien daaraan behoefte is ter verbetering van de kwaliteit. Indien waargenomen via de standaard CBS-vraagstelling gaat het om de volgende variabelen:

- Tijdens automatisch typeren: Beroep, VoornWzh (zie paragraaf 3.2).
- Tijdens handmatig typeren worden indien beschikbaar in aanvulling daarop nog gebruikt : Leiding, UitsLeid, DeelLeid, PersBeleid, StratBeleid, de SBI-code, Afl_HgstNivGev, N_LeidW, BedrsOmvZ (zie paragraaf 3.2 voor een toelichting op de variabelen).

Wanneer een externe partij een eigen beroepenvraagstelling heeft ontwikkeld en de daarmee verzamelde antwoorden wil laten typeren door het CBS is dat mogelijk onder bepaalde voorwaarden. Er moet gebruik zijn gemaakt van een open beroepenvraagstelling waarin ten minste gevraagd is naar het beroep van de respondent, en bij voorkeur ook naar de belangrijkste werkzaamheden. Overige relevante beroepenvariabelen kunnen alleen ingezet worden wanneer deze op dezelfde wijze zijn waargenomen als door het CBS, of daarnaar kunnen worden omgezet.

4.3 Output

De output die geleverd wordt is afhankelijk van de wens van de opdrachtgever met betrekking tot het aggregatieniveau van de ISCO 2008 of BRC 2014 en kwaliteit. Als gekozen is voor volledig automatische typering wordt de output geleverd op 4 digits (inclusief vangnetcodes) van de ISCO 2008. Slechts een klein deel, minder dan 5% is getypeerd met vangnetcodes, zie tabel 5. In aanvulling daarop wordt per typering de score geleverd die een maat is voor de waarschijnlijkheid dat de typering correct is. Met behulp van schakelschema's kunnen deze codes worden geschakeld naar de hogere aggregatieniveaus van de BRC 2014 en naar de ISCO 2008. In bijlages 2 en 3 wordt inzicht gegeven in de kwaliteit die dan per categorie gehaald wordt.

¹⁴ In deze stap worden indien gebruik is gemaakt van de vraagstelling over beroep van het CBS ook de variabelen over leidinggeven ingezet om records die in stap 1 en stap 2 zijn getypeerd met een ISCO-code van major group 1 Managers en die geen leiding geven door te sturen naar handmatig typeren. Dit wordt uitgevoerd ongeacht de score.

Tabel 5. Aandeel volledig automatisch getypeerd met ISCO 2008 codes op hoger aggregatieniveau met een vangnetcode of met unit group code of met beroep onbekend na doorlopen van stap 1 en stap 2 van procesvariant 2, EBB 2016 voor weging, 92 duizend records

Aggregatieniveau ISCO 2008	Aandeel t.o.v. totaal
	%
Vangnet major groups	2,4
Vangnet sub-major groups	1,3
Vangnetminor groups	0,3
unit groups	92,1
Onbekend	3,9
Totaal	100

Indien de opdrachtgever heeft aangegeven handmatig typeren te willen gebruiken om voldoende kwaliteit van alle categorieën binnen het gewenste aggregatieniveau van de beroepenindeling te behalen wordt het aggregatieniveau geleverd dat gevraagd is, of er kan daar via de SSB-werkwijze¹⁵ beschikking over worden verkregen.

4.4 Kwaliteit na inzetten handmatig typeren

Om de kwaliteit die gehaald wordt na de eerste twee automatische typeerstappen van procesvariant 2 meetbaar te maken wordt gebruik gemaakt van de EBB-jaarbestanden getypeerd via procesvariant 1 als vergelijkingsmateriaal. De omvang van deze dataset is voldoende om voor alle aggregatieniveaus en daaronder liggende groepen van de BRC 2014 en ISCO 2008 te kunnen bepalen welk deel nog handmatig getypeerd moet worden ter verbetering van de kwaliteit.

Uit een regressieanalyse werd informatie verkregen over de relatie tussen het aandeel met een ongelijke code tussen de beide processen en het aandeel met een onjuiste code van een handmatig gecontroleerde steekproef van 500 records, zie bijlage 1. Hieruit bleek dat om te voldoen aan het kwaliteitscriterium van maximaal 10% onjuiste typering per categorie van het 1^e tot en met 3^e aggregatieniveau van de BRC 2014 of ISCO 2008 het percentage met ongelijke code maximaal 10% mag zijn.

Tabel 6 laat voor een jaarbestand van de EBB zien dat bij afnemende score¹⁶ het aandeel met ongelijke code tussen de beide processen (volledig automatisch getypeerd via procesvariant 2 in vergelijking met getypeerd via procesvariant 1) toeneemt. Ook blijkt dat het aandeel met ongelijke code hoger is wanneer de typering in een meer gedetailleerd aggregatieniveau worden vergeleken. Over het geheel beschouwd is per aggregatieniveau de kwaliteit bij gebruik van de BRC 2014 iets beter dan de ISCO 2008. Bijvoorbeeld wanneer geaggregeerd naar de beroepsklassen van de BRC2014 is na volledige automatische typering 15% in een andere beroepsklasse getypeerd dan in procesvariant 1, maar geaggregeerd naar de major groups van de ISCO 2008 is 17% met een andere major group code getypeerd.

Uit deze tabel blijkt verder dat wanneer een scoredrempel van minimaal 40 wordt gehanteerd het aandeel met ongelijke code in de EBB in het 1^e tot en met 3^e aggregatieniveau van de BRC 2014 of ISCO 2008 11% of lager is. Deze drempelwaarde, een score van minimaal 40, is daarom in het proces gebruikt om het deel af te bakken dat nog handmatig getypeerd moet ter verbetering van de kwaliteit van het totaal van alle typering binnen

¹⁵ In het SSB worden ISCO 4-digite codes (inclusief vangnetcodes) opgeslagen, via referentieboeken kan het aggregatieniveau met voldoende kwaliteit worden aangekoppeld.

¹⁶ Tijdens het typeren van beroep worden alle typering door het programma Cascot voorzien van een score. De score is een maat voor de waarschijnlijkheid dat de toegekende code correct is. Hoe hoger de score hoe beter de kwaliteit.

het gewenste aggregatieniveau. Het deel met een score minder dan 40 wordt handmatig getypeerd wanneer binnen een bepaalde categorie het aandeel met een ongelijke code groter is dan 10%.

Tabel 6 Verdeling van de automatische typeringen over de scoreklassen, en per scoreklasse het aandeel met ongelijke code (volledig automatisch getypeerd via procesvariant 2 in vergelijking met getypeerd via procesvariant 1) per aggregatieniveau van de ISCO 2008 en BRC 2014, EBB 2016 ongewogen.

scoreklasse	# per scoreklasse	% tov totaal	BRC 2014			ISCO 2008			
			Klasse (2 digits)	Segment (3 digits)	Beroepsgroep (4 digits)	Major (1 dig)	Sub-major (2 dig)	Minor (3 dig)	Niveau (1 dig)
			% ongelijke code per scoreklasse			% ongelijke code per scoreklasse			
90-100	26034	28	1	1	1	1	1	1	1
80-89	7467	8	2	2	3	2	2	3	2
70-79	8479	9	3	4	4	3	4	5	3
60-69	9566	10	6	9	11	8	10	12	7
50-59	8695	9	13	19	23	16	20	24	13
40-49	11508	12	18	26	30	21	26	32	17
30-39	12699	14	36	55	64	39	58	67	34
20-29	3861	4	57	73	79	59	74	79	47
10-19	187	0	66	82	87	70	82	87	59
1-9	2	0	50	100	100	50	100	100	50
0 (beroep onbekend)	3577	4	76	76	76	81	81	81	81
Totaal	92075	100	15	20	23	17	21	24	14
40 of meer	71749	78	6	9	10	7	9	11	6
1 tot 40	16749	18	41	59	67	44	62	70	37
0 (beroep onbekend)	3577	4	76	76	76	81	81	81	81
Totaal	92075	100	15	20	23	17	21	24	14

De kwaliteit van output verkregen via procesvariant 2 na inzetten van handmatig typeren is verder onderzocht door middel van een pilot die in 2017 is uitgevoerd bij twee onderzoeken, het ICT-onderzoek (steekproef 3221 respondenten) en het AKO (steekproef 1318 respondenten). In tabel 7 staat de verdeling van de typeringen over de typeerstappen gepresenteerd voor de beide onderzoeken. Voor het ICT-onderzoek is het tweede aggregatieniveau, en voor het AKO het derde aggregatieniveau van de beroepenindeling het niveau dat met voldoende kwaliteit moet worden opgeleverd. In het ICT-onderzoek wordt dan nog 17% en in AKO nog 20% handmatig getypeerd. Ter vergelijking, in procesvariant 1 wordt 28% handmatig getypeerd.

Tabel 7 Verdeling over de typeerstappen indien handmatig typeren is ingezet. Voor ICT was outputeis het 2^e aggregatieniveau, voor AKO was de outputeis typeren het 3^e aggregatieniveau. Ongewogen data ICT ruim 3200 records en AKO ruim 1200 records.

Typeerstep	2e aggregatieniveau		3e aggregatieniveau	
	ICT		AKO	
		%		%
Automatisch				
1-beroep		71		63

2-beroep & werkzaamheden	12	16
3-handmatig typen	17	20
Totaal	100	100

Voor het ICT-onderzoek is onderzocht wat de kwaliteit is per beroepssegment (2^e aggregatieniveau) van de BRC 2014 na handmatig typen¹⁷. Ten opzichte van het totaal van alle typeringen heeft slechts 3% een onjuiste code gekregen. In 33 van de 41 beroepssegmenten is het aandeel met een onjuiste typering lager dan 10%. In 8 segmenten¹⁸ varieert het aandeel met een onjuiste code tussen de 10 en 15%. Hieruit blijkt dat bij onderzoeken met een veel kleinere steekproefomvang dan de EBB er rekening gehouden moet worden met schommelingen rondom de streefwaarde van maximaal 10% onjuiste typeringen, maar dat over het geheel genomen de kwaliteit goed is.

De kwaliteit van de automatische typeringen na doorlopen van de eerste twee typeersteps van procesvariant 2 wordt jaarlijks op basis van een geactualiseerd EBB-jaarbestand gemonitord door te controleren of er wijzigingen optreden in het aandeel met ongelijke code en het nodig is om aanpassingen te doen in de afbakening van handmatig te typen deel.

4.5 Kwaliteit na volledige automatisch typeringen

In bijlage 2 en 3 worden de tabellen gepresenteerd waarmee zicht op de kwaliteit per categorie in de verschillende aggregatieniveaus kan worden verkregen indien gekozen wordt voor volledig automatische typering. Er is hiervoor gebruik gemaakt van EBB verslagjaar 2016, een steekproef van 92 duizend records. Hieronder in tabel 8 wordt een fragment hiervan getoond en toegelicht hoe dit geïnterpreteerd moet worden.

In de eerste twee kolommen is voor de beide typeerprocessen per categorie van het 1^e, 2^e en 3^e aggregatieniveau van de en de BRC 2014 (bijlage 2) ISCO 2008 (bijlage 3) de relatieve omvang van de groep ten opzichte van het totaal getoond. De eerste laat het aandeel ten opzichte van het totaal zien wanneer data volledig automatisch is getypeerd, de tweede kolom laat het aandeel zien in het geval er is getypeerd volgens procesvariant 1. Dit geeft zicht op de mate waarin de relatieve omvang van een bepaalde groep binnen de hoger liggende aggregatieniveaus wijzigt na volledig automatische typering in vergelijking met het 4-digit typeerproces.

In de derde kolom is per volledig automatisch getypeerde categorie het aandeel berekend dat in procesvariant 1 een andere code heeft gekregen. Indien men handmatig typen wil inzetten om de kwaliteit te verbeteren binnen een bepaald aggregatieniveau moeten de records met een score minder dan 40 die behoren tot een categorie waarvan het aandeel ongelijk hoger is dan 10% handmatig worden getypeerd, zie paragraaf 4.4.

Per categorie van de BRC 2014 of ISCO 2008 is tevens een indicatie gegeven van de kwaliteit die verwacht kan worden indien ervoor is gekozen om data volledig automatisch te typen met het procesvariant 2. Deze indicatie is gebaseerd op een schatting van het aandeel met onjuiste code uitgaande van model C (zie bijlage 1) van de regressieanalyse waarin de relatie tussen het aandeel ongelijk en het aandeel onjuist is berekend. Het geschatte aandeel onjuist is verdeeld in 5 categorieën variërend van 'goed' met een geschat aandeel onjuist van 0-9% tot 'zeer slecht' met een geschat aandeel onjuist van meer dan 50%.

¹⁷ Interne nota, verslag pilot ICT AKO nieuw typeerproces beroep_v08012018. doc

¹⁸ Dit zijn 021-Auteurs en kunstenaars; 032-Vertegenwoordigers en inkoopers; 061-Overheidsambtenaren en -bestuurders; 075-Voedselverwerkende beroepen en overige ambachten; 076-Elektriciens en electronicamonteurs; 078-Hulpkrachten bouw en industrie; 082-Vakspecialisten ICT en 092-Hulpkrachten landbouw, Interne notitie Resultaten pilot ICT en AKO, typeerproces beroep typen hoger aggregatieniveau, 08-01-2018

Tabel 8 (fragment bijlage 2) De verdeling van de typering over de beroepsklassen, -segmenten en -groepen van de BRC 2014 volgens procesvariant 1 en volledig automatisch getypeerd volgens procesvariant 2, het aandeel per volledig automatisch getypeerde categorie met een andere code in procesvariant 1 en een indicatie van de kwaliteit, EBB 2016 ongewogen totaal 92 duizend records.

BRC2014, Beroepsklasse, -segment, -groep	Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie	Kwaliteitsindicatie bij volledig automatisch typeren o.b.v. schatting aandeel onjuist				
	Aandeel tov totaal	Aandeel tov totaal		Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
	%	%	%					
01 Pedagogische beroepen	7,16%	7,09%	9,6%	x				
011 Docenten	4,87%	4,94%	8,1%	x				
Docenten hoger onderwijs en hoogleraren	0,69%	0,60%	33,7%				x	
Docenten beroepsgerichte vakken secundair onderwijs	0,34%	0,40%	38,5%					x
Docenten algemene vakken secundair onderwijs	1,11%	1,26%	18,0%			x		
>>>voor resterende regels, zie bijlage 2<<<								
Totaal	100%	100%						

Indien men volledig automatisch getypeerde data wil gebruiken voor analyse- en publicatiedoeleinden wordt geadviseerd om rekening te houden met de kwaliteit alvorens uitspraken te doen met betrekking tot individuele categorieën van de ISCO 2008 of BRC 2014 en een verwijzing op te nemen naar documentatie waarin nagezocht kan worden wat de verwachte kwaliteit is.

In bijlage 4 wordt door middel van kruistabellen inzicht gegeven in hoe volledig automatische typering zijn verdeeld over de major groups van de ISCO 2008 of de beroepsklassen van de BRC 2014 na typering met procesvariant 1. Dit geeft inzicht in waar records met een ongelijke code volgens procesvariant 1 ingedeeld zouden worden. Zo blijkt bijvoorbeeld dat 90,4% van records die na volledige automatische typering zijn getypeerd in beroepsklasse 1 'Pedagogische beroepen' ook in die beroepsklasse zijn getypeerd in procesvariant 1. De resterende records zijn verdeeld over alle andere beroepsklassen, maar voornamelijk in beroepsklasse 10 'Zorg en welzijn' en 4 'Bedrijfseconomische en administratieve beroepen'. Indien daaraan behoefte is kunnen deze tabellen ook geleverd worden voor het 2^e of 3^e aggregatieniveau van de ISCO 2008 of BRC 2014.

Bijlage 1. Meten kwaliteit volledig automatisch typen

Om zicht te krijgen op de kwaliteit per categorie van de beroepenindelingen zijn de codes die in de EBB verslagjaar 2015 zijn toegekend in procesvariant 1 vergeleken met de codes die zijn toegekend als er volledig automatisch wordt getypeerd via procesvariant 2.

Een random selectie van 500 records met codes volgens procesvariant 1 en na volledig automatische typering is handmatig gecontroleerd. Per record is beoordeeld of de code die is toegekend in het volledig automatische typeerproces correct is, en gecorrigeerd indien niet correct.

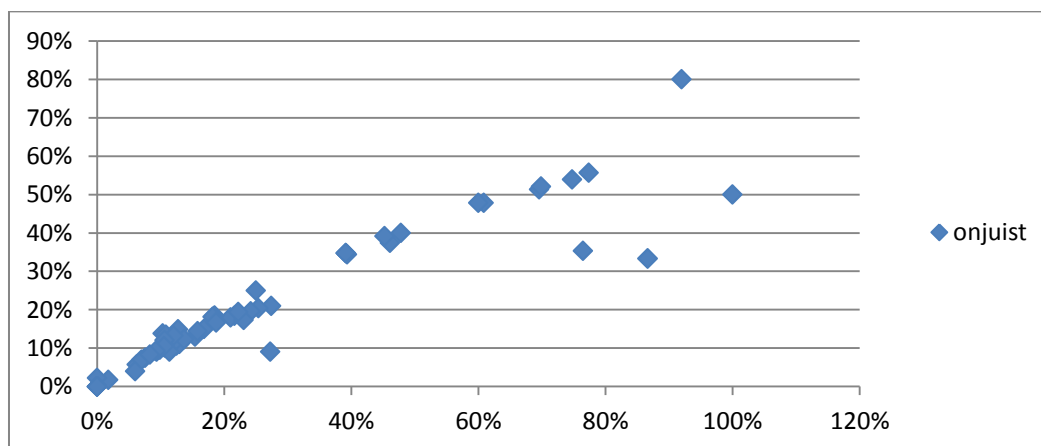
Om zicht te krijgen op het verband tussen het percentage met een code die ongelijk is aan het huidige proces en het percentage met een onjuiste code is de gecontroleerde selectie verdeeld in verschillende groepen, zie tabel 1a. In totaal worden er 59 verschillende groepen onderscheiden. Het verband tussen het % ongelijk en % onjuist is in figuur 1 uitgezet in spreidingsdiagram. Zowel uit tabel 1a als uit figuur 1 blijkt een verband tussen aandeel ongelijk en aandeel onjuist, per groep ligt het aandeel met een ongelijke en onjuiste code dicht bij elkaar als de verschillen in de coderingen tussen de beide datasets kleiner zijn. Dit blijkt bijvoorbeeld in tabel 1a uit de groepen getypeerd met score 40 en meer versus getypeerd met een score minder dan 40.

Omdat uit de figuur een lineair verband blijkt is het precieze verband tussen deze variabelen berekend door middel van een lineaire regressieanalyse die is toegepast op alle groepen (model A), de groepen met een minimale omvang van 10 waarnemingen (model B), en groepen met een minimale omvang van 25 waarnemingen (model C), zie tabel 1b. Het percentage onjuist is als afhankelijke, en het percentage ongelijk als onafhankelijke variabele gebruikt.

Uit tabel 1b blijkt dat alle modellen significant zijn, en dat model C, dus waarbij alleen groepen met meer dan 25 waarnemingen zijn meegenomen het meest betrouwbaar is. Uitgaande van de constante en coëfficiënt van model C mag het percentage ongelijk niet hoger dan 10% om een percentage onjuist van 10% op te leveren, volgens de modellen A en B mag het percentage ongelijk niet hoger dan 9% zijn.

Op basis van deze bevindingen is in de analyses naar de kwaliteit van het volledig automatisch typeerproces de stelregel gehanteerd dat als binnen een categorie het aandeel met ongelijke code maximaal 10% is, het percentage met een onjuiste code dan niet hoger dan 10% zal zijn en daarmee van voldoende kwaliteit.

Figuur 1. Het verband tussen het percentage met ongelijke code (x-as) en het percentage met onjuiste code (y-as) van de groepen gepresenteerd in tabel 1a, steekproef 500 records EBB 2015.



Tabel 1a. Typeringen volledig automatisch getypeerd in de verschillende aggregatieniveaus van de ISCO en BRC en naar scoreklasse met per groep het aandeel met ongelijke code aan procesvariant 1 en aandeel met onjuiste code, steekproef van 500 records uit de EBB 2015. Oranje gearceerde groepen zijn weggelaten in model B, blauw en oranje gearceerde groepen in model C.

	Aantal per groep	% ongelijk	% onjuist
ISCO 2008			
Totaal			
Unit groups (4 digits)	500	27,4%	21,0%
Minor groups (3 digits)	500	25,4%	20,4%
Sub-major groups (2 digits)	500	21,6%	18,4%
Major groups (1 digit)	500	16,8%	15,0%
Niveau	500	15,4%	13,0%
Score 40 en meer			
Unit groups (4 digits)	385	12,5%	10,6%
Minor group (3 digits)	385	10,6%	10,4%
Sub-major group (2 digits)	385	9,9%	9,6%
Major groups (1 digit)	385	7,5%	7,5%
Niveau	385	6,2%	5,7%
Score minder dan 40			
Unit groups (4 digits)	115	77,4%	55,7%
Minor group (3 digits)	115	74,8%	53,9%
Sub-major group (2 digits)	115	60,9%	47,8%
Major groups (1 digit)	115	47,8%	40,0%
Niveau	115	46,1%	37,4%
Major groups			
00~Armed forces occupations	1	0,0%	0,0%
01~Managers	23	39,1%	34,8%
02~Professionals	136	11,8%	11,0%
03~Technicians and associate professionals	65	18,5%	18,5%
04~Clerical support workers	44	11,4%	9,1%
05~Service and sales workers	116	10,3%	13,8%
06~Skilled agricultural, forestry and fishery workers	6	0,0%	0,0%
07~Craft and related trades workers	47	12,8%	14,9%
08~Plant and machine operators, and assemblers	11	27,3%	9,1%
09~Elementary occupations	36	22,2%	19,4%
Beroep onbekend	15	86,7%	33,3%
BRC 2014			
Totaal			
Beroepsgroep	500	24,2%	19,6%
Beroepssegment	500	21,0%	18,0%
Beroepsklasse	500	15,8%	14,4%
Score 40 en meer			
Beroepsgroep	385	10,6%	10,1%
Beroepssegment	385	9,4%	9,1%
Beroepsklasse	385	7,0%	7,0%
Score minder dan 40			
Beroepsgroep	115	69,6%	51,3%
Beroepssegment	115	60,0%	47,8%
Beroepsklasse	115	45,2%	39,1%
Beroepsklassen			
01-Pedagogische beroepen	37	10,8%	13,5%
02-Creatieve en taalkundige beroepen	16	25,0%	25,0%

03-Commerciële beroepen	73	13,7%	12,3%
04-Bedrijfseconomische en administratieve beroepen	83	12,0%	13,3%
05-Managers	23	39,1%	34,8%
06-Openbaar bestuur, veiligheid en juridische beroepen	11	18,2%	18,2%
07-Technische beroepen	66	10,6%	12,1%
08-ICT beroepen	19	10,5%	10,5%
09-Agrarische beroepen	7	0,0%	0,0%
10-Zorg en welzijn beroepen	76	9,2%	9,2%
11-Dienstverlenende beroepen	48	18,8%	16,7%
12-Transport en logistiek beroepen	24	8,3%	8,3%
13-Overig	17	76,5%	35,3%
Totaal naar scoreklasse			
score100	10	0,0%	0,0%
score90-99	113	1,8%	1,8%
score80-89	45	0,0%	2,2%
score70-79	50	6,0%	4,0%
score60-69	54	13,0%	11,1%
score50-59	52	23,1%	17,3%
score40-49	61	39,3%	34,4%
score30-39	73	69,9%	52,1%
score20-29	25	92,0%	80,0%
score10-19	2	100,0%	50,0%
score0	15	86,7%	33,3%

Tabel 1b. Samenvatting regressieanalyse van relatie tussen percentage ongelijk (x-waarde) en percentage onjuist (y-waarde): model A alle groepen, model B groepen met minimale omvang 10 waarnemingen, model C groepen met minimale omvang 25 waarnemingen.

	N		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	x-waarde als y=10%
			B	Std. Error	Beta			
Model A	59	(Constant)	4,614	1,304		3,539	0,001	9,1
		ongelijk	0,589	0,034	0,918	17,477	0,000	
Model B	55	(Constant)	4,643	1,381		3,361	0,001	8,9
		ongelijk	0,605	0,037	0,915	16,546	0,000	
Model C	44	(Constant)	2,305	0,526		4,380	0,000	10,2
		ongelijk	0,751	0,015	0,991	49,298	0,000	

Bijlage 2. Kwaliteit typeerproces volledig automatisch BRC 2014

De verdeling van de typering over de beroepsklassen, -segmenten en -groepen van de BRC 2014 volgens procesvariant 1 en volledig automatisch getypeerd volgens procesvariant 2, het aandeel per volledig automatisch getypeerde categorie met een andere code in procesvariant 1 en een indicatie van de kwaliteit, EBB 2016 ongewogen totaal 92 duizend records.

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ¹⁹	Kwaliteitsindicatie bij volledig automatisch typeren o.b.v. schatting aandeel onjuist ²⁰				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20- 29% onjuist)	Slecht (30- 49% onjuist)	Zeer slecht (>49% onjuist)
BRC2014, Beroepsklasse, -segment, -groep		%	%	%					
01	Pedagogische beroepen	7,16%	7,09%	9,6%	x				
011	Docenten	4,87%	4,94%	8,1%	x				
0111	Docenten hoger onderwijs en hoogleraren	0,69%	0,60%	33,7%			x		
0112	Docenten beroepsgerichte vakken secundair onderwijs	0,34%	0,40%	38,5%				x	
0113	Docenten algemene vakken secundair onderwijs	1,11%	1,26%	18,0%		x			
0114	Leerkrachten basisonderwijs	1,90%	1,90%	15,7%		x			
0115	Onderwijskundigen en overige docenten	0,82%	0,78%	31,2%				x	
012	Sportinstructeurs	0,51%	0,56%	19,4%		x			
0121	Sportinstructeurs	0,51%	0,56%	19,4%		x			
013	Leidsters kinderopvang en onderwijsassistenten	1,79%	1,58%	21,8%		x			
0131	Leidsters kinderopvang en onderwijsassistenten	1,79%	1,58%	21,8%		x			
02	Creatieve en taalkundige beroepen	2,30%	2,07%	22,7%		x			

¹⁹ Handmatig typeren binnen een bepaalde categorie wordt aanbevolen wanneer het aandeel met ongelijke code hoger is dan 10%, onderliggende categorieën met voldoende kwaliteit hoeven niet handmatig te worden getypeerd. Bijvoorbeeld beroepsgroep 0214 Beeldend kunstenaars hoeft niet getypeerd te worden, maar de overige beroepsgroepen binnen segment 021 Auteurs en kunstenaars wel om de kwaliteit van het hele segment te verbeteren.

²⁰ Aandeel onjuist is berekend door uit te gaan van in bijlage 1 toegelichte regressieanalyse naar het verband tussen percentage met ongelijke code en met onjuiste code, model C.

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ¹⁹	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²⁰				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20- 29% onjuist)	Slecht (30- 49% onjuist)	Zeer slecht (>49% onjuist)
BRC2014, Beroepsklasse, -segment, -groep		%	%	%					
021	Auteurs en kunstenaars	1,39%	1,27%	22,8%		x			
0211	Bibliothecarissen en conservatoren	0,10%	0,09%	29,2%				x	
0212	Auteurs en taalkundigen	0,33%	0,35%	13,2%		x			
0213	Journalisten	0,30%	0,28%	21,6%		x			
0214	Beeldend kunstenaars	0,12%	0,13%	2,8%	x				
0215	Uitvoerend kunstenaars	0,55%	0,43%	35,9%				x	
022	Vakspecialisten op artistiek en cultureel gebied	0,91%	0,79%	25,4%				x	
0221	Grafisch vormgevers en productontwerpers	0,60%	0,52%	26,1%				x	
0222	Fotografen en interieurontwerpers	0,30%	0,27%	25,4%				x	
03	Commerciële beroepen	10,50%	11,87%	10,6%		x			
031	Adviseurs marketing, public relations en sales	1,42%	1,65%	26,4%					x
0311	Adviseurs marketing, public relations en sales	1,42%	1,65%	26,4%					x
032	Vertegenwoordigers en inkopers	1,31%	1,64%	34,3%					x
0321	Vertegenwoordigers en inkopers	1,31%	1,64%	34,3%					x
033	Verkopers	7,76%	8,57%	10,2%	x				
0331	Winkeliers en teamleiders detailhandel	0,81%	1,32%	37,9%					x
0332	Verkoopmedewerkers detailhandel	4,09%	4,51%	9,1%	x				
0333	Kassamedewerkers	1,38%	1,28%	9,6%	x				
0334	Callcentermedewerkers outbound en overige verkopers	1,48%	1,45%	21,5%			x		
04	Bedrijfseconomische en administratieve beroepen	18,21%	18,71%	12,1%		x			
041	Specialisten bedrijfsbeheer en administratie	5,17%	5,32%	21,5%		x			
0411	Accountants	1,26%	1,06%	25,4%				x	
0412	Financieel specialisten en economen	0,89%	0,83%	29,6%				x	
0413	Bedrijfskundigen en organisatieadviseurs	1,32%	1,57%	32,7%				x	

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ¹⁹	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²⁰				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20- 29% onjuist)	Slecht (30- 49% onjuist)	Zeer slecht (>49% onjuist)
BRC2014, Beroepsklasse, -segment, -groep		%	%	%					
0414	Beleidsadviseurs	0,82%	0,72%	22,7%		x			
0415	Specialisten personeels- en loopbaanontwikkeling	0,88%	1,14%	13,9%		x			
042	Vakspecialisten bedrijfsbeheer en administratie	3,42%	3,74%	17,9%		x			
0421	Boekhouders	1,11%	1,22%	12,7%		x			
0422	Zakelijke dienstverleners	1,01%	1,11%	26,6%				x	
0423	Directiesecretaresses	1,30%	1,41%	16,7%		x			
043	Administratief personeel	9,62%	9,65%	13,2%		x			
0431	Administratief medewerkers	3,35%	3,55%	21,3%		x			
0432	Secretaresses	0,72%	0,63%	21,1%		x			
0433	Receptionisten en telefonisten	2,16%	1,93%	21,9%		x			
0434	Boekhoudkundig medewerkers	1,69%	1,60%	22,2%		x			
0435	Transportplanners en logistiek medewerkers	1,70%	1,94%	19,4%		x			
05	Managers	6,20%	5,23%	36,0%				x	
051	Algemeen directeuren	0,47%	0,88%	36,7%				x	
0511	Algemeen directeuren	0,47%	0,88%	36,7%				x	
052	Managers op administratief en commercieel gebied	2,21%	1,48%	45,4%					x
0521	Managers zakelijke en administratieve dienstverlening	1,14%	0,91%	33,6%				x	
0522	Managers verkoop en marketing	1,07%	0,57%	59,1%					x
053	Managers productie en gespecialiseerde dienstverlening	1,39%	1,88%	35,2%				x	
0531	Managers productie	0,25%	0,55%	36,9%					x
0532	Managers logistiek	0,20%	0,21%	45,4%					x
0533	Managers ICT	0,30%	0,23%	46,0%					x
0534	Managers zorginstellingen	0,20%	0,33%	23,4%		x			
0535	Managers onderwijs	0,16%	0,22%	20,8%		x			

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ¹⁹	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²⁰				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20- 29% onjuist)	Slecht (30- 49% onjuist)	Zeer slecht (>49% onjuist)
BRC2014, Beroepsklasse, -segment, -groep		%	%	%					
0536	Managers gespecialiseerde dienstverlening	0,28%	0,34%	41,9%				x	
054	Managers horeca, detailhandel en overige diensten	0,26%	0,68%	35,9%			x		
0541	Managers horeca	0,12%	0,19%	27,2%			x		
0542	Managers detail- en groothandel	0,06%	0,37%	15,5%		x			
0543	Managers commerciële en persoonlijke dienstverlening	0,07%	0,12%	69,2%					x
055	Managers zonder nadere differentiatie	1,87%	0,31%	96,8%					x
0551	Managers zonder nadere differentiatie	1,87%	0,31%	96,8%					x
06	Openbaar bestuur, veiligheid en juridische beroepen	3,03%	3,29%	16,1%		x			
061	Overheidsambtenaren en -bestuurders	0,94%	1,08%	30,7%			x		
0611	Overheidsbestuurders	0,24%	0,41%	38,7%				x	
0612	Overheidsambtenaren	0,69%	0,66%	30,7%			x		
062	Juristen	0,67%	0,76%	7,2%	x				
0621	Juristen	0,67%	0,76%	7,2%	x				
063	Beveiligingswerkers	1,42%	1,45%	15,5%		x			
0631	Politie-inspecteurs	0,12%	0,12%	14,0%		x			
0632	Politie en brandweer	0,41%	0,41%	15,2%		x			
0633	Beveiligingspersoneel	0,64%	0,60%	17,0%		x			
0634	Militaire beroepen	0,26%	0,32%	17,4%		x			
07	Technische beroepen	12,27%	12,85%	13,8%		x			
071	Ingenieurs en onderzoekers wis-, natuur- en technische wetenschappen	2,19%	2,25%	24,2%			x		
0711	Biologen en natuurwetenschappers	0,41%	0,33%	38,0%				x	
0712	Ingenieurs (geen elektrotechniek)	1,22%	1,45%	23,5%		x			
0713	Elektrotechnisch ingenieurs	0,13%	0,13%	29,1%			x		

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ¹⁹	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²⁰				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20- 29% onjuist)	Slecht (30- 49% onjuist)	Zeer slecht (>49% onjuist)
BRC2014, Beroepsklasse, -segment, -groep		%	%	%					
0714	Architecten	0,43%	0,35%	31,8%			x		
072	Vakspecialisten natuur en techniek	1,78%	1,98%	28,1%			x		
0721	Technici bouwkunde en natuur	1,01%	1,08%	27,9%			x		
0722	Productieleiders industrie en bouw	0,51%	0,65%	28,4%			x		
0723	Procesoperators	0,25%	0,24%	33,2%			x		
073	Bouwarbeiders	2,53%	2,61%	12,6%		x			
0731	Bouwarbeiders ruwbouw	0,61%	0,65%	24,2%				x	
0732	Timmerlieden	0,79%	0,80%	9,6%	x				
0733	Bouwarbeiders afbouw	0,36%	0,37%	14,3%		x			
0734	Loodgieters en pijpfitters	0,31%	0,36%	17,7%		x			
0735	Schilders en metaalspuiters	0,46%	0,43%	12,2%		x			
074	Metaalarbeiders, machinemonteurs	1,91%	1,92%	15,6%		x			
0741	Metaalbewerkers en constructiewerkers	0,47%	0,47%	20,6%		x			
0742	Lassers en plaatwerkers	0,37%	0,36%	12,4%		x			
0743	Automonteurs	0,61%	0,61%	16,1%		x			
0744	Machinemonteurs	0,46%	0,48%	22,6%		x			
075	Voedselverwerkende beroepen en overige ambachten	1,36%	1,33%	25,3%				x	
0751	Slagers	0,21%	0,18%	26,3%				x	
0752	Bakkers	0,23%	0,22%	11,4%		x			
0753	Productcontroleurs	0,20%	0,24%	53,8%					x
0754	Meubelmakers, kleermakers en stoffeerders	0,40%	0,40%	17,8%		x			
0755	Medewerkers drukkerij en kunstnijverheid	0,32%	0,29%	28,5%			x		
076	Elektriciens en electronicamonteurs	0,79%	0,87%	24,1%				x	
0761	Elektriciens en electronicamonteurs	0,79%	0,87%	24,1%				x	

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ¹⁹	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²⁰				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20- 29% onjuist)	Slecht (30- 49% onjuist)	Zeer slecht (>49% onjuist)
BRC2014, Beroepsklasse, -segment, -groep		%	%	%					
077	Productiemachinebedieners en assemblagemedewerkers	0,95%	1,08%	36,9%			x		
0771	Productiemachinebedieners	0,74%	0,79%	39,5%				x	
0772	Assemblagemedewerkers	0,21%	0,29%	33,3%			x		
078	Hulpkrachten bouw en industrie	0,78%	0,81%	21,2%		x			
0781	Hulpkrachten bouw en industrie	0,78%	0,81%	21,2%		x			
08	ICT beroepen	3,67%	3,79%	14,8%		x			
081	Specialisten ICT	3,07%	3,16%	14,4%		x			
0811	Software- en applicatieontwikkelaars	2,28%	2,39%	16,7%		x			
0812	Databank- en netwerkspecialisten	0,79%	0,77%	12,4%		x			
082	Vakspecialisten ICT	0,60%	0,63%	35,2%			x		
0821	Gebruikersondersteuning ICT	0,44%	0,49%	34,2%			x		
0822	Radio- en televisietechnici	0,17%	0,14%	39,1%				x	
09	Agrarische beroepen	1,88%	2,11%	9,0%	x				
091	Tuinders, akkerbouwers en veetelers	1,47%	1,74%	9,4%	x				
0911	Land- en bosbouwers	0,23%	0,33%	32,9%				x	
0912	Hoveniers, tuinders en kwekers	0,79%	0,86%	12,9%		x			
0913	Veetelers	0,46%	0,55%	15,1%		x			
092	Hulpkrachten landbouw	0,40%	0,37%	24,8%				x	
0921	Hulpkrachten landbouw	0,40%	0,37%	24,8%				x	
10	Zorg en welzijn beroepen	13,39%	13,12%	9,7%	x				
101	Artsen, therapeuten en gespecialiseerd verpleegkundigen	3,55%	3,33%	16,9%		x			
1011	Artsen	1,26%	1,16%	15,1%		x			
1012	Gespecialiseerd verpleegkundigen	1,54%	1,38%	25,2%				x	
1013	Fysiotherapeuten	0,75%	0,80%	8,8%	x				

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ¹⁹	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²⁰				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20- 29% onjuist)	Slecht (30- 49% onjuist)	Zeer slecht (>49% onjuist)
BRC2014, Beroepsklasse, -segment, -groep		%	%	%					
102	Specialisten op maatschappelijk gebied	1,49%	1,63%	15,4%		x			
1021	Maatschappelijk werkers	0,74%	0,84%	20,1%		x			
1022	Psychologen en sociologen	0,75%	0,80%	11,7%		x			
103	Vakspecialisten gezondheidszorg	2,69%	2,83%	15,6%		x			
1031	Laboranten	0,27%	0,26%	23,0%		x			
1032	Apothekersassistenten	0,23%	0,26%	1,9%	x				
1033	Verpleegkundigen (mbo)	0,78%	0,91%	18,6%		x			
1034	Medisch praktijkassistenten	0,80%	0,84%	3,8%	x				
1035	Medisch vakspecialisten	0,61%	0,55%	32,1%				x	
104	Sociaal werkers, groeps- en woonbegeleiders	2,82%	2,55%	23,9%				x	
1041	Sociaal werkers, groeps- en woonbegeleiders	2,82%	2,55%	23,9%				x	
105	Verzorgenden	2,83%	2,78%	11,9%		x			
1051	Verzorgenden	2,83%	2,78%	11,9%		x			
11	Dienstverlenende beroepen	9,69%	10,01%	7,1%	x				
111	Medewerkers persoonlijke dienstverlening	5,50%	5,79%	8,8%	x				
1111	Reisbegeleiders	0,25%	0,25%	12,1%		x			
1112	Koks	0,68%	0,78%	4,8%	x				
1113	Kelners en barpersoneel	2,91%	3,11%	5,7%	x				
1114	Kappers en schoonheidsspecialisten	0,72%	0,78%	3,0%	x				
1115	Conciërges en teamleiders schoonmaak	0,48%	0,50%	32,7%				x	
1116	Verleners van overige persoonlijke diensten	0,46%	0,39%	28,6%				x	
112	Schoonmakers en keukenhulpen	4,18%	4,22%	8,0%	x				
1121	Schoonmakers	3,08%	3,07%	8,9%	x				
1122	Keukenhulpen	1,11%	1,15%	7,0%	x				

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ¹⁹	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²⁰				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20- 29% onjuist)	Slecht (30- 49% onjuist)	Zeer slecht (>49% onjuist)
BRC2014, Beroepsklasse, -segment, -groep		%	%	%					
12	Transport en logistiek beroepen	7,59%	7,80%	5,6%	x				
121	Bestuurders voertuigen en bedieners mobiele machines	2,96%	3,11%	6,0%	x				
1211	Dekofficieren en piloten	0,26%	0,26%	25,2%				x	
1212	Chauffeurs auto's, taxi's en bestelwagens	0,70%	0,70%	12,6%		x			
1213	Buschauffeurs en trambestuurders	0,25%	0,22%	16,7%		x			
1214	Vrachtwagenchauffeurs	1,13%	1,19%	10,9%		x			
1215	Bedieners mobiele machines	0,62%	0,73%	6,3%	x				
122	Hulpkrachten transport en logistiek	4,63%	4,69%	7,1%	x				
1221	Laders, lossers en vakkenvullers	3,48%	3,52%	6,5%	x				
1222	Vuilnisophalers en dagbladenbezorgers	1,15%	1,17%	10,1%	x				
13	Overig²¹	4,12%	2,07%	75,2%					x
131	Overig	4,12%	2,07%	75,2%					x
1311	Overig	4,12%	2,07%	75,2%					x
Totaal		100,00%	100,00%						

²¹ Overig omvat personen waarvan beroep onbekend of niet in te delen is

Bijlage 3. Kwaliteit typeerproces volledig automatisch ISCO 2008

De verdeling van de typeringen over de ISCO 2008 major groups, sub-major groups en minor groups volgens procesvariant 1 en volledig automatisch getypeerd volgens procesvariant 2, het aandeel per volledig automatisch getypeerde categorie met een andere code in procesvariant 1 en, een indicatie van de kwaliteit, EBB 2016 ongewogen totaal 92 duizend records.

ISCO 2008 major group, sub-major group, minor group	Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typeren o.b.v. schatting aandeel onjuist ²³				
	Aandeel tov totaal	Aandeel tov totaal	%	Goed (0-9% onjuist)	Voldoende (10-19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
	%	%	%					
0 Armed forces occupations	0,26%	0,32%	17,4%	X				
00 Armed forces occupations (not further defined)	0,01%	0,10%	50,0%				X	
000 Armed forces occupations (not further defined)	0,01%	0,10%	50,0%				X	
01 Commissioned armed forces officers	0,02%	0,05%	82,6%					X
011 Commissioned armed forces officers	0,02%	0,05%	82,6%					X
02 Non-commissioned armed forces officers	0,06%	0,07%	60,4%				X	
021 Non-commissioned armed forces officers	0,06%	0,07%	60,4%				X	
03 Armed forces occupations, other ranks	0,17%	0,09%	75,8%					X
031 Armed forces occupations, other ranks	0,17%	0,09%	75,8%					X
1 Managers	6,44%	5,64%	34,9%			X		
10 Managers (not further defined)	1,86%	0,31%	96,8%					X
100 Managers (not further defined)	1,86%	0,31%	96,8%					X

²² Handmatig typeren binnen een bepaalde categorie wordt aanbevolen wanneer het aandeel met ongelijke code hoger is dan 10%, onderliggende categorieën met voldoende kwaliteit hoeven niet handmatig te worden getypeerd.

²³ Aandeel onjuist is berekend door uit te gaan van in bijlage 1 toegelichte regressieanalyse naar het verband tussen percentage met ongelijke code en met onjuiste code, model C.

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²³				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0- 9% onjuist)	Voldoen de (10- 19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
ISCO 2008 major group, sub-major group, minor group		%	%	%					
11	Chief executives, senior officials and legislators	0,71%	1,29%	35,1%			X		
111	Legislators and senior officials	0,24%	0,41%	38,7%				X	
112	Managing directors and chief executives	0,47%	0,88%	36,7%				X	
12	Administrative and commercial managers	2,22%	1,48%	45,7%				X	
120	Administrative and commercial managers (not further defined)	0,01%	0,91%	100,0%					X
121	Business services and administration managers	1,14%	0,57%	33,6%			X		
122	Sales, marketing and development managers	1,07%	1,88%	59,1%				X	
13	Production and specialised services managers	1,39%	0,00%	35,2%			X		
132	Manufacturing, mining, construction, and distribution managers	0,45%	0,75%	40,4%				X	
133	Information and communications technology service managers	0,30%	0,23%	46,0%				X	
134	Professional services managers	0,64%	0,89%	30,4%			X		
14	Hospitality, retail and other services managers	0,26%	0,68%	35,9%			X		
141	Hotel and restaurant managers	0,12%	0,19%	27,2%			X		
142	Retail and wholesale trade managers	0,06%	0,37%	15,5%		X			
143	Other services managers	0,07%	0,12%	69,2%					X
2	Professionals	24,43%	24,85%	12,2%		X			
20	Professionals (not further defined)	0,11%	0,36%	12,4%		X			
200	Professionals (not further defined)	0,11%	0,36%	12,4%		X			
21	Science and engineering professionals	2,69%	2,42%	26,3%			X		
210	Science and engineering professionals (not further defined)	0,00%	0,00%	100,0%					X
211	Physical and earth science professionals	0,08%	0,07%	40,8%				X	
212	Mathematicians, actuaries and statisticians	0,08%	0,06%	31,4%			X		

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²³				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0- 9% onjuist)	Voldoen de (10- 19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
ISCO 2008 major group, sub-major group, minor group		%	%	%					
213	Life science professionals	0,25%	0,20%	39,5%				X	
214	Engineering professionals (excluding electrotechnology)	1,11%	1,09%	26,7%			X		
215	Electrotechnology engineers	0,13%	0,13%	29,1%			X		
216	Architects, planners, surveyors and designers	1,04%	0,87%	27,8%			X		
22	Health professionals	3,55%	3,33%	17,0%		X			
221	Medical doctors	1,00%	0,93%	14,7%		X			
222	Nursing and midwifery professionals	1,54%	1,38%	25,2%			X		
223	Traditional and complementary medicine professionals	0,04%	0,05%	47,1%				X	
224	Paramedical practitioners	0,01%	0,01%	27,3%			X		
225	Veterinarians	0,06%	0,06%	7,4%	X				
226	Other health professionals	0,90%	0,91%	9,9%	X				
23	Teaching professionals	4,87%	4,94%	8,1%	X				
230	Teaching professionals (not further defined)	0,31%	0,23%	76,8%					X
231	University and higher education teachers	0,69%	0,60%	33,7%			X		
232	Vocational education teachers	0,34%	0,40%	38,5%				X	
233	Secondary education teachers	1,11%	1,26%	18,0%		X			
234	Primary school and early childhood teachers	1,36%	1,45%	8,2%	X				
235	Other teaching professionals	1,05%	1,00%	30,8%			X		
24	Business and administration professionals	6,54%	6,92%	21,2%		X			
240	Business and administration professionals (not further defined)	0,01%	0,00%	100,0%					X
241	Finance professionals	2,09%	1,84%	26,7%			X		
242	Administration professionals	3,01%	3,43%	21,7%		X			
243	Sales, marketing and public relations professionals	1,42%	1,65%	26,4%			X		

ISCO 2008 major group, sub-major group, minor group	Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²³				
	Aandeel tov totaal	Aandeel tov totaal	%	Goed (0- 9% onjuist)	Voldoen de (10- 19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
25	Information and communications technology professionals	3,07%	3,16%	14,4%	X			
	Information and communications technology professionals (not further defined)	0,10%	0,09%	67,7%				X
	250 Software and applications developers and analysts	2,18%	2,29%	16,0%	X			
	252 Database and network professionals	0,79%	0,77%	12,4%	X			
26	Legal, social and cultural professionals	3,62%	3,72%	16,5%	X			
	261 Legal professionals	0,67%	0,76%	7,2%	X			
	262 Librarians, archivists and curators	0,10%	0,09%	29,2%			X	
	263 Social and religious professionals	1,56%	1,69%	15,7%	X			
	264 Authors, journalists and linguists	0,63%	0,63%	14,8%	X			
	265 Creative and performing artists	0,66%	0,56%	30,2%			X	
3	Technicians and associate professionals	14,62%	15,37%	19,7%	X			
31	Science and engineering associate professionals	2,03%	2,24%	27,5%				X
	311 Physical and engineering science technicians	0,95%	0,99%	29,3%			X	
	312 Mining, manufacturing and construction supervisors	0,51%	0,65%	28,4%			X	
	313 Process control technicians	0,25%	0,24%	33,2%			X	
	314 Life science technicians and related associate professionals	0,06%	0,09%	15,8%	X			
	315 Ship and aircraft controllers and technicians	0,26%	0,26%	25,2%			X	
32	Health associate professionals	2,69%	2,83%	15,6%	X			
	321 Medical and pharmaceutical technicians	0,57%	0,58%	16,3%	X			
	322 Nursing and midwifery associate professionals	0,78%	0,91%	18,6%	X			
	323 Traditional and complementary medicine associate professionals		0,01%					
	324 Veterinary technicians and assistants	0,03%	0,03%	0,0%	X			
	325 Other health associate professionals	1,31%	1,30%	15,2%			X	

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typeren o.b.v. schatting aandeel onjuist ²³				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0- 9% onjuist)	Voldoen de (10- 19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
ISCO 2008 major group, sub-major group, minor group		%	%	%					
33	Business and administration associate professionals	5,54%	6,16%	21,6%	X				
331	Financial and mathematical associate professionals	1,11%	1,22%	12,7%	X				
332	Sales and purchasing agents and brokers	1,31%	1,64%	34,3%			X		
333	Business services agents	1,01%	1,11%	26,6%			X		
334	Administrative and specialised secretaries	1,30%	1,41%	16,7%	X				
335	Regulatory government associate professionals	0,81%	0,78%	28,0%			X		
34	Legal, social, cultural and related associate professionals	3,74%	3,51%	22,2%	X				
340	Legal, social, cultural and related associate professionals (not further defined)	0,00%	0,00%	100,0%					X
341	Legal, social and religious associate professionals	2,82%	2,55%	23,9%			X		
342	Sports and fitness workers	0,51%	0,56%	19,4%	X				
343	Artistic, cultural and culinary associate professionals	0,41%	0,39%	20,5%	X				
35	Information and communications technicians	0,60%	0,63%	35,2%			X		
351	Information and communications technology operations and user support technicians	0,44%	0,44%	34,9%			X		
352	Telecommunications and broadcasting technicians	0,17%	0,14%	39,1%				X	
4	Clerical support workers	9,62%	9,65%	13,2%	X				
40	Clerical support workers (not further defined)	0,20%	0,16%	44,3%				X	
400	Clerical support workers (not further defined)	0,20%	0,16%	44,3%				X	
41	General and keyboard clerks	2,05%	2,52%	18,6%	X				
410	General and keyboard clerks (not further defined)	0,48%	0,36%	57,4%				X	
411	General office clerks	0,84%	1,53%	24,8%			X		
412	Secretaries (general)	0,64%	0,55%	17,7%	X				
413	Keyboard operators	0,08%	0,08%	49,3%				X	

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²³				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0- 9% onjuist)	Voldoen de (10- 19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
ISCO 2008 major group, sub-major group, minor group		%	%	%					
42	Customer services clerks	2,34%	2,08%	22,1%		X			
421	Tellers, money collectors and related clerks	0,18%	0,15%	28,0%			X		
422	Client information workers	2,16%	1,93%	21,9%		X			
43	Numerical and material recording clerks	3,39%	3,54%	20,5%		X			
431	Numerical clerks	1,69%	1,60%	22,2%		X			
432	Material-recording and transport clerks	1,70%	1,94%	19,4%		X			
44	Other clerical support workers	1,64%	1,35%	27,2%				X	
441	Other clerical support workers	1,64%	1,35%	27,2%				X	
5	Service and sales workers	18,83%	19,62%	10,2%		X			
50	Service and sales workers (not further defined)	0,00%	0,00%	100,0%					X
500	Service and sales workers (not further defined)	0,00%	0,00%	100,0%					X
51	Personal service workers	5,39%	5,67%	8,9%	X				
510	Personal service workers (not further defined)	0,01%	0,00%	100,0%					X
511	Travel attendants, conductors and guides	0,25%	0,25%	12,1%		X			
512	Cooks	0,56%	0,66%	4,6%	X				
513	Waiters and bartenders	2,91%	3,11%	5,7%	X				
514	Hairdressers, beauticians and related workers	0,72%	0,78%	3,0%	X				
515	Building and housekeeping supervisors	0,48%	0,50%	32,7%				X	
516	Other personal services workers	0,46%	0,39%	27,2%				X	
52	Sales workers	7,76%	8,57%	10,2%	X				
520	Sales workers (not further defined)	0,00%	0,00%	100,0%					X
521	Street and market salespersons	0,04%	0,08%	27,5%				X	
522	Shop salespersons	4,90%	5,84%	11,4%		X			
523	Cashiers and ticket clerks	1,38%	1,28%	9,6%	X				

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typeren o.b.v. schatting aandeel onjuist ²³				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0- 9% onjuist)	Voldoen de (10- 19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
ISCO 2008 major group, sub-major group, minor group		%	%	%					
524	Other sales workers	1,44%	1,37%	21,5%		X			
53	Personal care workers	4,62%	4,36%	15,4%		X			
530	Personal care workers (not further defined)	0,00%	0,00%	100,0%					X
531	Child care workers and teachers' aides	1,79%	1,58%	21,8%		X			
532	Personal care workers in health services	2,83%	2,78%	11,9%		X			
54	Protective services workers	1,05%	1,02%	15,6%		X			
541	Protective services workers	1,05%	1,02%	15,6%		X			
6	Skilled agricultural, forestry and fishery workers	1,47%	1,74%	9,4%	X				
61	Market-oriented skilled agricultural workers	1,45%	1,72%	9,1%	X				
611	Market gardeners and crop growers	0,91%	1,12%	11,1%		X			
612	Animal producers	0,44%	0,53%	14,3%		X			
613	Mixed crop and animal producers	0,10%	0,05%	73,4%					X
62	Market-oriented skilled forestry, fishery and hunting worker	0,02%	0,02%	33,3%			X		
621	Forestry and related workers	0,00%	0,01%	0,0%	X				
622	Fishery workers, hunters and trappers	0,02%	0,01%	35,3%			X		
7	Craft and related trades workers	6,81%	7,32%	12,1%		X			
70	Craft and related trades workers (not further defined)	0,22%	0,60%	59,2%					X
700	Craft and related trades workers (not further defined)	0,22%	0,60%	59,2%					X
71	Building and related trades workers, excluding electricians	2,53%	2,61%	12,6%		X			
711	Building frame and related trades workers	1,40%	1,45%	14,4%		X			
712	Building finishers and related trades workers	0,66%	0,73%	15,4%		X			
713	Painters, building structure cleaners and related trades workers	0,46%	0,43%	12,2%		X			
72	Metal, machinery and related trades workers	1,91%	1,92%	15,6%		X			

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typeren o.b.v. schatting aandeel onjuist ²³				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0- 9% onjuist)	Voldoen de (10- 19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
ISCO 2008 major group, sub-major group, minor group		%	%	%					
720	Metal, machinery and related trades workers (not further defined)	0,00%	0,00%	100,0%					X
721	Sheet and structural metal workers, moulders and welders, and related workers	0,53%	0,55%	12,1%		X			
722	Blacksmiths, toolmakers and related trades workers	0,31%	0,29%	25,8%			X		
723	Machinery mechanics and repairers	1,07%	1,09%	17,7%		X			
73	Handicraft and printing workers	0,32%	0,29%	28,5%			X		
731	Handicraft workers	0,14%	0,11%	34,1%			X		
732	Printing trades workers	0,18%	0,18%	24,3%			X		
74	Electrical and electronic trades workers	0,79%	0,87%	24,1%			X		
741	Electrical equipment installers and repairers	0,70%	0,76%	24,9%			X		
742	Electronics and telecommunications installers and repairers	0,09%	0,11%	32,5%			X		
75	r Food processing, wood working, garment and other craft and	1,04%	1,04%	24,6%			X		
751	Food processing and related trades workers	0,52%	0,49%	24,9%			X		
752	Wood treaters, cabinet-makers and related trades workers	0,20%	0,21%	11,8%		X			
753	Garment and related trades workers	0,20%	0,19%	24,5%			X		
754	Other craft and related workers	0,12%	0,15%	51,4%				X	
8	Plant and machine operators, and assemblers	3,65%	3,93%	12,5%		X			
81	Stationary plant and machine operators	0,74%	0,79%	39,5%				X	
810	Stationary plant and machine operators (not further defined)	0,33%	0,19%	75,1%					X
811	Mining and mineral processing plant operators	0,04%	0,03%	58,3%				X	
812	Metal processing and finishing plant operators	0,03%	0,05%	41,9%				X	
813	Chemical and photographic products plant and machine operators	0,04%	0,07%	38,9%				X	

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typeren o.b.v. schatting aandeel onjuist ²³				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0- 9% onjuist)	Voldoen de (10- 19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
ISCO 2008 major group, sub-major group, minor group		%	%	%					
814	Rubber, plastic and paper products machine operators	0,05%	0,12%	55,3%				X	
815	Textile, fur and leather products machine operators	0,10%	0,10%	52,2%				X	
816	Food and related products machine operators	0,05%	0,13%	65,2%					X
817	Wood processing and papermaking plant operators	0,02%	0,04%	57,1%				X	
818	Other stationary plant and machine operators	0,07%	0,05%	65,2%					X
82	Assemblers	0,21%	0,29%	33,3%			X		
821	Assemblers	0,21%	0,29%	33,3%			X		
83	Drivers and mobile plant operators	2,70%	2,85%	4,4%	X				
830	Drivers and mobile plant operators (not further defined)	0,03%	0,39%	96,8%					X
831	Locomotive engine drivers and related workers	0,04%	0,06%	10,0%	X				
832	Car, van and motorcycle drivers	0,70%	0,70%	12,6%		X			
833	Heavy truck and bus drivers	1,35%	1,03%	28,0%			X		
834	Mobile plant operators	0,55%	0,64%	5,3%	X				
835	Ships' deck crews and related workers	0,03%	0,03%	20,0%		X			
9	Elementary occupations	9,99%	10,09%	8,3%	X				
91	Cleaners and helpers	3,08%	3,07%	8,9%	X				
910	Cleaners and helpers (not further defined)	0,02%	0,00%	100,0%					X
911	Domestic, hotel and office cleaners and helpers	2,81%	2,84%	8,6%	X				
912	Vehicle, window, laundry and other hand cleaning workers	0,25%	0,23%	21,9%		X			
92	Agricultural, forestry and fishery labourers	0,40%	0,37%	24,8%			X		
921	Agricultural, forestry and fishery labourers	0,40%	0,37%	24,8%			X		
93	Labourers in mining, construction, manufacturing and transpo	4,26%	4,33%	8,7%	X				
931	Mining and construction labourers	0,11%	0,13%	22,4%		X			
932	Manufacturing labourers	0,67%	0,68%	21,2%		X			

		Procesvariant 2 (volledig automatisch)	Procesvariant 1	Aandeel ongelijke code per volledig automatisch getypeerde categorie ²²	Kwaliteitsindicatie bij volledig automatisch typen o.b.v. schatting aandeel onjuist ²³				
		Aandeel tov totaal	Aandeel tov totaal		Goed (0- 9% onjuist)	Voldoen de (10- 19% onjuist)	Matig (20-29% onjuist)	Slecht (30-49% onjuist)	Zeer slecht (>49% onjuist)
ISCO 2008 major group, sub-major group, minor group		%	%	%					
933	Transport and storage labourers	3,48%	3,52%	6,5%	X				
94	Food preparation assistants	1,11%	1,15%	7,0%	X				
941	Food preparation assistants	1,11%	1,15%	7,0%	X				
95	Street and related sales and service workers	0,01%	0,01%	61,5%				X	
951	Street and related service workers	0,00%	0,00%	0,0%	X				
952	Street vendors (excluding food)	0,01%	0,01%	66,7%					X
96	Refuse workers and other elementary workers	1,13%	1,16%	9,5%	X				
961	Refuse workers	0,08%	0,12%	22,4%		X			
962	Other elementary workers	1,05%	1,05%	8,6%	X				
99	Onbekend	3,88%	1,47%	81,4%					X
99	Weet niet of onbekend	3,88%	1,47%	81,4%					X
999	Weet niet of onbekend	3,88%	1,47%	81,4%					X
		100,00%	0,00%						

Bijlage 4. Kruistabellen 1^e aggregatieniveau BRC 2014 en ISCO 2008

Tabel 1. Verdeling per volledig automatisch getypeerde beroepsklasse van de BRC 2014 over beroepsklassen getypeerd volgens procesvariant 1, EBB 2016, ruim 92 duizend records ongewogen

	Procesvariant 1													Totaal
	01	02	03	04	05	06	07	08	09	10	11	12	13	
Volledig automatisch, procesvariant 2	%													
01-Pedagogische beroepen	90,4	0,2	0,3	1,7	0,7	0,4	0,8	0,2	0,1	3,4	0,9	0,1	0,8	100
02-Creatieve en taalkundige beroepen	1,6	77,3	3,5	4,1	2,0	0,4	5,2	2,5	0,2	0,9	0,5	0,2	1,7	100
03-Commerciële beroepen	0,1	0,3	89,4	3,1	1,6	0,5	1,1	0,4	0,3	0,4	1,3	1,0	0,6	100
04-Bedrijfseconomische en administratieve beroepen	0,4	0,3	2,9	87,9	1,4	1,1	2,2	0,8	0,2	1,0	0,6	0,5	0,9	100
05-Managers	0,6	0,5	14,1	8,2	64,0	1,4	3,5	2,0	1,1	0,8	1,9	0,6	1,4	100
06-Openbaar bestuur, veiligheid en juridische beroepen	0,5	0,2	1,3	5,6	1,3	83,9	2,2	0,7	0,1	1,9	0,6	0,7	1,1	100
07-Technische beroepen	0,4	0,4	1,7	2,5	0,8	0,5	86,2	0,8	0,7	0,9	1,3	1,2	2,8	100
08-ICT beroepen	0,3	0,8	1,3	4,2	1,3	0,6	4,0	85,2	0,0	0,7	0,3	0,2	1,1	100
09-Agrarische beroepen	0,2	0,1	1,7	1,4	0,3	0,2	2,2	0,0	91,0	1,0	0,8	0,8	0,4	100
10-Zorg en welzijn beroepen	1,9	0,1	0,4	2,1	0,7	0,7	1,5	0,3	0,1	90,3	0,8	0,2	1,0	100
11-Dienstverlenende beroepen	0,4	0,1	1,1	1,0	0,5	0,4	1,0	0,1	0,5	1,0	92,9	0,4	0,6	100
12-Transport en logistiek beroepen	0,1	0,1	1,2	0,9	0,1	0,5	1,3	0,1	0,1	0,2	0,7	94,4	0,4	100
13-Overig (inclusief onbekend)	1,9	1,4	9,7	13,9	9,7	2,4	17,2	2,9	2,8	4,4	5,4	3,5	24,8	100
Totaal	7,1	2,1	11,9	18,7	5,2	3,3	12,8	3,8	2,1	13,1	10,0	7,8	2,1	100

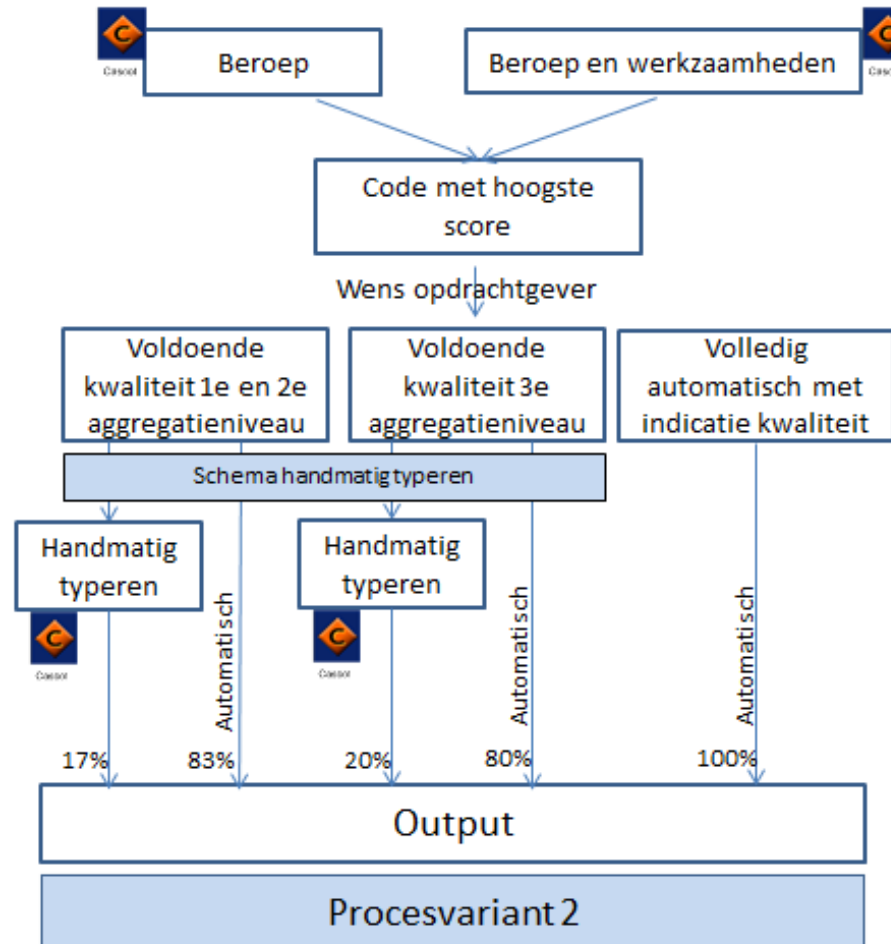
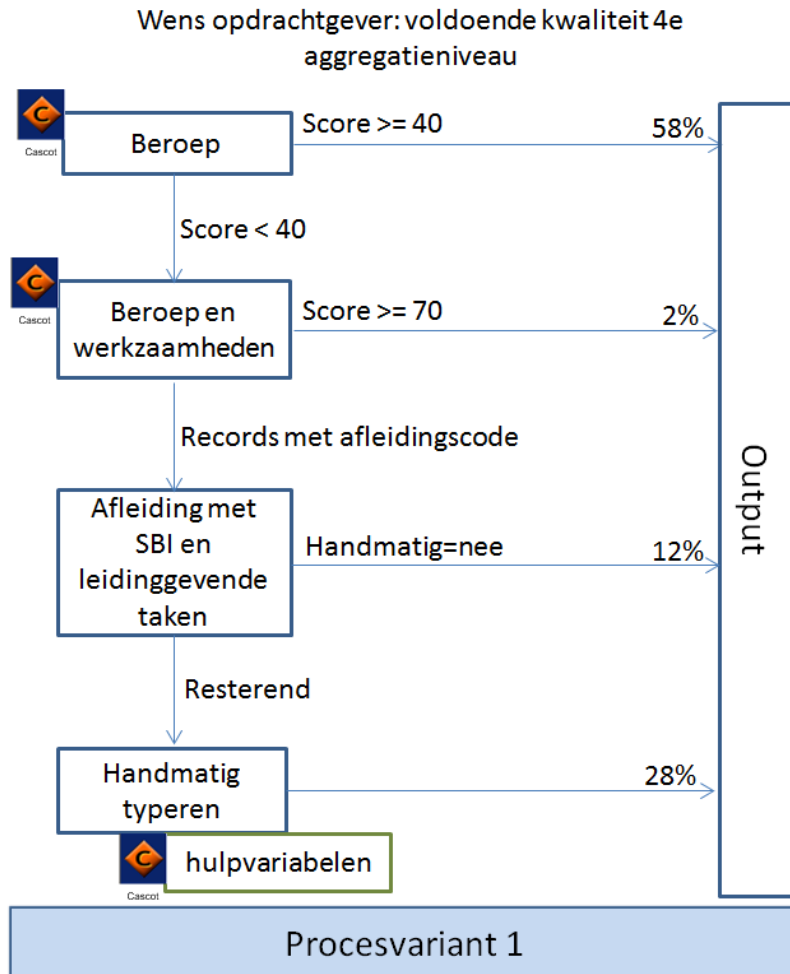
Tabel 2. Verdeling per volledig automatisch getypeerde ISCO 2008 major group over major groups getypeerd volgens procesvariant 1, EBB 2016, ruim 92 duizend records ongewogen

	Procesvariant 1											99 totaal
	0	1	2	3	4	5	6	7	8	9		
Volledig automatisch, procesvariant 2	%											
0-Armed forces occupations	82,6	0,9	2,1	3,4	0,4	6,8	0,0	1,3	0,9	0,0	1,7	100
1-Managers	0,1	65,1	11,9	10,7	1,3	6,2	1,1	1,5	0,6	0,3	1,3	100
2-Professionals	0,1	1,9	87,8	4,9	1,2	1,4	0,2	0,9	0,3	0,3	0,9	100
3-Technicians and associate professionals	0,3	1,4	7,7	80,3	2,6	3,6	0,2	1,4	0,8	0,4	1,2	100
4-Clerical support workers	0,1	0,6	2,3	4,5	86,8	3,1	0,1	0,7	0,4	0,8	0,7	100
5-Service and sales workers	0,0	1,2	2,0	3,2	1,0	89,8	0,2	0,4	0,2	1,4	0,4	100

6-Skilled agricultural, forestry and fishery workers	0,1	0,4	0,7	1,4	1,1	2,0	90,6	1,0	0,7	1,7	0,3	100
7-Craft and related trades workers	0,1	0,4	1,6	2,1	0,6	2,4	0,4	87,9	2,2	1,6	0,6	100
8-Plant and machine operators, and assemblers	0,1	0,1	0,7	1,7	0,9	1,9	0,4	3,5	87,5	2,8	0,6	100
9-Elementary occupations	0,0	0,1	0,5	0,8	0,8	3,3	0,7	0,8	0,9	91,7	0,4	100
99-Onbekend	0,1	11,3	17,9	10,1	4,6	13,4	2,4	11,3	4,8	5,3	18,6	100
Totaal	0,3	5,6	24,9	15,4	9,7	19,6	1,7	7,3	3,9	10,1	1,5	100

Bijlage 5. Schematisch overzicht en praktijkvoorbeelden

Schematische weergave van de beide procesvarianten, percentages gebaseerd op tabel 2 en tabel 7.



Praktijkvoorbeelden

Onderstaande voorbeelden zijn bedoeld ter illustratie van een aantal mogelijke manieren waarop antwoorden op beroep en werkzaamheden de beide processen doorlopen, het is niet de bedoeling om een compleet overzicht te geven.

Voorbeeld 1

Beroep: interieurverzorger

Werkzaamheden: schoonmaken vh interieur

Procesverloop variant 1

Het record doorloopt stap 1 en krijgt met score 0 een code voor beroep onbekend. De typering met onbekend is via conclusion-regels van Cascot geregeld. Na combinatie met werkzaamheden krijgt de record met score 57 een match met index-entry 'interieur verzorger huishouden' ISCO-code 9111 Schoonmakers in huishoudens. De score is echter lager dan 70, dit is in stap 2 de drempelwaarde die wordt gehanteerd. Er is in stap 1 en in stap 2 geen afleidingscode toegekend waardoor de record niet stap 3 doorloopt en vervolgens in stap 4 handmatig wordt getypeerd. De typeur beslist in dit geval of gegeven de aanvullende hulpvariabelen een typering met een vangnetcode 9110 van toepassing is, of dat de persoon schoonmaakt bij particuliere huishoudens (9111) of bij hotels, kantoren (9112).

Procesverloop variant 2

Het record doorloopt stap 1, en krijgt met score 69 een match met index-entry 'interieur verzorger huishouden', ISCO-code 9111 Schoonmakers in huishoudens. De record doorloopt vervolgens na combinatie met de werkzaamheden stap 2 en krijgt met score 56 een match met index-entry 'interieur verzorger huishouden', ISCO-code 9111. De score van de typering in stap 1 is hoger dan die in stap 2, de typering van stap 1 wordt overgenomen met bijbehorende score in de output. Omdat de score hoger is dan 40 hoeft niet meer handmatig te worden getypeerd, ook als de opdrachtgever aangegeven handmatig te typeren te willen inzetten om de kwaliteit te verbeteren. In dit geval is het weliswaar onbekend of de persoon in een particulier huishouden, of bij een bedrijf of instelling schoonmaakt, de typering is echter wel correct op het 3^e aggregatieniveau van de ISCO (minor group 911 Domestic, hotel and office cleaners and helpers) of BRC 2014 (beroepsgroep 1121 Schoonmakers).

Voorbeeld 2

Beroep: voorbereidend med

Werkzaamheden: groentes voorbereiden en afwas

Variant 1

Het record doorloopt stap 1 en krijgt met score 33 een match met index entry 'box medewerker' ISCO-code 4321 'logistiek medewerkers, medewerkers materiaal-, voorraadplanning en -beheer'. Deze score is lager dan 40, het record gaat daarom door naar stap 2. Na toevoegen van de werkzaamheden in stap 2 wordt een match met

index entry 'afwasmedewerker' met score 61 ISCO-code 9412 'Keukenhulpen'. Er is geen afleidingscode toegekend in de eerste twee typeerstappen, en de score is in de tweede stap lager dan 70. Het record daarom door naar stap 4 waar deze handmatig wordt getypeerd met dezelfde code als vastgelegd in stap 2.

Variant 2

Het proces verloopt hetzelfde als in variant 1, maar nu wordt de code overgenomen in de output die in stap 2 is vastgelegd, de score is immers hoger dan die in stap 1 is vastgelegd en ook hoger dan 40. Er hoeft niet meer handmatig te worden getypeerd.

Voorbeeld 3

Beroep: Accountmanager

Werkzaamheden: Inzicht verkrijgen in vraag van nieuwe en bestaande klanten

Variant 1

Het record doorloopt stap 1 en krijgt met score 99 een afleidingscode toegekend. Dit wordt via de defaultcoding regels van Cascot geregeld. In de tweede stap worden de werkzaamheden gecombineerd met het beroep en wordt met score 26 een match gevonden met index entry 'accountmanager pensioenen' ISCO-code 3321 'Verzekeringsagenten'. Dit is lager dan 70, en omdat in de eerste stap een afleidingscode is toegekend worden de afleidingsregels doorlopen in stap 3. Gegeven de SBI-code 6202 'Advisering op het gebied van informatietechnologie' waar deze persoon werkzaam is wordt in deze stap de ISCO-code 2434 'vertegenwoordigers informatie- en communicatietechnologie' toegekend. Deze code komt in de output terecht.

Variant 2

Het record doorloopt stap 1 en krijgt met score 38 een match met index-entry 'accountmanager retail' ISCO-code 3322 'Vertegenwoordigers, accountmanagers retail en exportmanagers' gevonden. Deze score en gevonden index-entry zijn geregeld via de cascot-regels. In de tweede stap wordt na combinatie met de werkzaamheden met score 26 een match gevonden met index entry 'accountmanager pensioenen' ISCO-code 3321 'verzekeringsagenten'. Dit is lager dan de score van de eerste stap, de code vastgelegd in de eerste stap komt dan ook in de output terecht. Als er gekozen is om handmatig typeren in te zetten om de kwaliteit te verbeteren wordt deze record doorgestuurd naar handmatig typeren, want zoals blijkt uit bijlage 3 is binnen minor group 332 het aandeel met ongelijke code aan procesvariant 1 hoger dan 10%, ook in de beroepsgroep 0321 'Vertegenwoordigers en inkopers' van de BRC 2014 waar deze code toe behoort is het aandeel hoger dan 10%. Tijdens handmatig typeren wordt op basis van de informatie over de SBI ISCO-code 2434 toegekend.

Nb. Deze aanpak wordt voor meerdere beroepen gekozen die in procesvariant 1 via de afleidingsregels op basis van beschikbare hulpvariabelen naar een ISCO-code worden afgeleid. In procesvariant 2 worden deze beroepen naar de meest waarschijnlijke code gestuurd uitgaande van de beroepsomschrijving, maar wordt de score ook (indien nodig) gestuurd naar een waarde lager dan 40 zodat deze records afgebakend kunnen worden om handmatig te typeren wanneer er behoefte is aan hogere kwaliteit.

Voorbeeld 4

Beroep: Opticien

Werkzaamheden: Ogen meten en advies geven

Procesvariant 1:

Het record krijgt op basis van de omschrijving van het beroep in stap 1 met score 97 een match met indexentry 'opticien' de ISCO-code 3254 Opticien. De score is hoger dan de drempelwaarde en gaat naar de output.

Procesvariant 2:

Het record krijgt op basis van de omschrijving van het beroep in stap 1 met score 97 een match met indexentry 'opticien' de ISCO-code 3254 Opticien. In de tweede stap worden de werkzaamheden gecombineerd met het beroep en wordt met score 52 een match gevonden met dezelfde index entry 'opticien' en de ISCO-code 3254 Opticien. De typering met de hoogste score van de eerste stap wordt overgenomen in de output.

