

Hoe kwaliteitsvol is een autonome, AI-gedreven clustering van competenties als taxonomie?

Synthese van de kwalitatieve expertenbeoordelingen.

Wouter Dossche, Sarah Vansteenkiste

Aanbevolen referentie bij gebruik:

Dossche, W., & Vansteenkiste, S. (2026). *Hoe kwaliteitsvol is een autonome, AI-gedreven clustering van competenties als taxonomie? Synthese van de kwalitatieve expertenbeoordelingen*. Steunpunt Werk. Werk.Rapport 2026 nr.1.

ISBN 9789088731440

Samenvatting kwalitatieve bevraging (clusterkwaliteit)

Dit rapport vat de bevraging samen naar de kwaliteit van drie verschillende methodes voor het clusteren van vacatureskills voor de beroepen accountant en industrieel elektromecaniciën. De bevraging werd afgenomen bij de zes leden van het VDAB Competent 2-team en twee leden van Steunpunt Werk. Het opzet van de bevraging was te bepalen welke volledig autonome, data-gedreven methode het meest geschikt is als uitgangspunt voor het groeperen van gelijkaardige skills, en welke relatie tussen skills het best aansluit bij menselijke intuïtie voor verdere verfijning. De clustering werd uitgevoerd op het meest granulaire hiërarchische niveau, aangezien de kwaliteit van elementaire skill-groeperingen een noodzakelijke voorwaarde vormt voor betekenisvolle aggregatie op hogere niveaus. De data omvatte zowel de AI-geëxtraheerde skills uit vacatureteksten als de op de vacatures vermelde Competent 1-competenties. In een tweede luik werd de data-gedreven clustering expliciet afgezet tegen de expert-gedreven Competent 2.0-taxonomie, met als doel te evalueren of en hoe beide benaderingen elkaar kunnen aanvullen.

De volgende preprocessingstappen werden doorlopen:

- Toewijzen van het Competent 2 beroep aan de vacatures o.b.v. een modelvoorspelling (AI-team), de skills in de vacature en de vacaturetitel
- Data cleaning voor het verwijderen van testvacatures, vacatures voor IBO/opleidingen/stages/..., en vacatures met weinig informatie
- Deduplicatie van vacatures
- Skills werden gefilterd zodanig dat 97% van alle skillvermeldingen gecapteerd werden
- Een skillfilter voor het verwijderen van zelden voorkomende skills, vreemd ogende/foutieve/irrelevante skills, soft skills
- Toevoeging van skills die minder frequent voorkomen, maar een sterke recente stijging in de vraag ernaar kennen

De volgende methodes kwamen aan bod in het onderzoek:

- **Methode A** hield uitsluitend rekening met het samen voorkomen van skills in vacatures (co-occurrence patronen). Om dit te modelleren werd Leiden community detection met Markov Stability toegepast op een skill co-occurrence knowledge graph. De graph bestond uit één node per skill, waarbij relaties werden bepaald via het Ochiai-coëfficiënt dat co-occurrence normaliseert voor skillfrequentie. Het Ochiai-coëfficiënt werd getransformeerd met een tijdsvervalfunctie zodat recentere observaties een pak sterker doorwegen dan oudere observaties.
- **Methode B** hield zowel rekening met de co-occurrence patronen als de semantische gelijkenis tussen skills. Hiervoor werd een Graph Neural Network toegepast op dezelfde knowledge graph als methode A, waarbij de skillteksten via een taalmodel werden omgezet in variabelen (tekst embeddings) en toegevoegd aan de nodes.
- **Methode C** hield uitsluitend rekening met de semantische gelijkenis tussen skills. Om dit te modelleren werd hetzelfde taalmodel als methode B gebruikt om de teksten van de skills om te zetten in variabelen. Vervolgens werd een klassiek clusteralgoritme (k-means) toegepast.

Luik 1: Vergelijking van de data-gedreven methodes

De bevroagde items werden geselecteerd om verschillende, complementaire kwaliteitsaspecten van clustering te evalueren. Idealiter bestaat een clustering uit groepen met hoge interne samenhang en lage externe gelijkenis, zodat observaties binnen een cluster sterk op elkaar lijken terwijl clusters onderling duidelijk van elkaar verschillen. Wat als interne samenhang geldt, kan echter verschillend worden ingevuld, bijvoorbeeld op basis van inhoudelijke (semantische) gelijkenis of op basis van functionele samenhang, waarbij elementen gezamenlijk voorkomen of samen vereist zijn binnen eenzelfde context. De bevroagde items operationaliseerden deze principes door zowel semantische coherentie als functionele (taak-gerelateerde) coherentie, naast onderscheidbaarheid, ruis en globale kwaliteit, te evalueren.

De onderstaande tabel geeft weer hoe de methodes gemiddeld scoorden op de verschillende items. De Likert-schaal antwoorden werden omgezet naar een schaal van -3 (Helemaal oneens), 0 (Neutraal), 3 (Helemaal eens). De antwoorden per item voor elk beroep kunnen in de appendix aan het eind van dit rapport gevonden worden.

	Accountant			Industrieel elektromecanici			Total		
	A	B	C	A	B	C	A	B	C
Semantische Clustercoherentie	-1.62 (1.19)	-1.25 (1.28)	0.88 (1.13)	-2.12 (0.99)	-1.50 (1.20)	0.88 (0.99)	-1.88 (1.09)	-1.38 (1.20)	0.88 (1.02)
Taakgerelateerde Coherentie	-1.75 (1.49)	-1.75 (0.71)	-0.25 (1.49)	-2.00 (1.07)	-1.62 (1.06)	-0.38 (1.51)	-1.88 (1.26)	-1.69 (0.87)	-0.31 (1.45)
Cluster Onderscheidbaarheid	-1.62 (1.19)	-1.38 (0.92)	0.38 (1.92)	-2.12 (0.99)	-1.75 (1.04)	0.25 (0.89)	-1.88 (1.09)	-1.56 (0.96)	0.31 (1.45)
Cluster Redundantie	-2.00 (0.76)	-1.38 (1.06)	0.25 (1.67)	-2.12 (0.99)	-1.88 (0.99)	0.12 (1.46)	-2.06 (0.85)	-1.62 (1.02)	0.19 (1.52)
Overzichtelijkheid/ bruikbaarheid	-1.50 (1.20)	-1.25 (1.04)	-0.12 (1.96)	-2.38 (1.06)	-1.62 (1.19)	-0.25 (1.16)	-1.94 (1.18)	-1.44 (1.09)	-0.19 (1.56)
Algemene kwaliteit	-1.88 (0.99)	-1.50 (0.93)	0.12 (1.73)	-2.00 (1.07)	-1.62 (1.19)	0.12 (1.13)	-1.94 (1.00)	-1.56 (1.03)	0.12 (1.41)

Gemiddelde score (Standaarddeviatie) op item over beoordelaars heen. De beste score wordt aangeduid in het vet.

Over alle metriecken heen kwam een zeer stabiel en consistent patroon naar voren over beide beroepen en beoordelingsgroepen. Methode A scoorde systematisch het zwakst, met lage beoordelingen voor zowel semantische als taakgerelateerde coherentie en een hoge waargenomen redundantie. Daarbij bleek de sterk ongelijke verdeling van de clustergroottes, met vele singletons (clusters bestaande uit één skill) en enkele grote clusters, ook ten koste te gaan van de kwaliteit. In tegenstelling tot de hypothese bleek deze methode globaal genomen onvoldoende adequaat in het capteren van skills die gezamenlijk vereist zijn voor een bepaald takenpakket. De beoordelaars gaven meerdere voorbeelden van clusters die onlogisch overkwamen. Een post-hoc analyse leerde dat de skills in enkele van deze clusters ook daadwerkelijk zelden samen voorkomen in vacatures, waarbij we vaststelden dat een skill kan worden toegewezen aan een cluster wanneer deze een sterk verwantschap toont met één of enkele andere skill(s) in de cluster, ondanks dat de skill voorts weinig verband hield met de overige skills. Dergelijke misleidende co-occurrence patronen werden voornamelijk vastgesteld bij weinig frequente skills. Een alternatieve verklaring die we niet kunnen uitsluiten is dat de aanwezigheid van de C1-competenties, welke heel frequent voorkwamen en vaak gezamenlijk gevraagd werden volgens het standaardvoorstel, aanzienlijke ruis opleverde die de overige co-occurrence patronen overheerste. In sommige

gevallen zagen de beoordelaars geen logica in de clusters, hoewel de cluster skills bevatte die wel degelijk frequent samen voorkomen in vacatures. Dit suggereert dat valide co-occurrence patronen niet noodzakelijk overeenstemmen met menselijke intuïtie. Daarbij gaven beoordelaars ook aan dat zij verwachtten dat talen, rijbewijzen of bepaalde software elk onder één cluster vallen, terwijl deze skills zelden allen gezamenlijk gevraagd worden. Een verdere vaststelling is dat deze methode ook de tendens had om generieke of heel frequente skills samen te hokken die voorts weinig gemeenschappelijks deelden met elkaar.

Methode B vertoonde een gematigde verbetering ten opzichte van methode A, maar bleef inconsistent en vertoonde dezelfde kinderziekten. De clusters werden occasioneel als taakgerelateerd gezien, maar bleken te heterogeen om als concreet taakgerelateerd te functioneren. Anderzijds stelden we vast dat de semantische coherentie verbeterde zonder concessies te doen aan de taakgerelateerde coherentie t.a.v. methode A. Deze hogere semantische coherentie vertaalde zich gedeeltelijk naar de bruikbaarheid en kwaliteitsbeoordelingen toe. Een beoordelaar merkte tevens op dat sommige clusters inhoudelijk coherent leken te zijn, terwijl andere clusters eerder een samenvatting van een deel van een functie leken te zijn. Dit oppert het idee dat het gezamenlijk rekening houden met semantiek en co-occurrence patronen ook inconsistenties in de groeperingen kan voortbrengen.

Methode C scoorde over de volledige lijn beter en werd consistent erkend als beste methode, al expliciteerden beoordelaars dat deze keuze relatief was en geenszins als eindproduct gebruikt kon worden. Methode C bleek enigszins de artefacten van co-occurrence (linken met zeldzame skills, dominante frequenties) te reduceren en leverde clusters die beoordelaars eerder als inhoudelijk herkenbaar achten. Conform de verwachting resulteerde methode C in clusters die eerder goed scoren op semantische coherentie, maar dit vertaalde zich niet noodzakelijk door naar functionele of taakgerichte samenhang (e.g., “graafkranen”, “productiemachines”, “zaagmachines” werden ondergebracht in een cluster gerelateerd aan machines, maar zullen zoals aangegeven door een beoordelaar zelden samen gevraagd worden). Hoewel de clusters volgens beoordelaars grotendeels inhoudelijke samenhang vertoonden, bleven de kwaliteitsbeoordelingen eerder laag. Dit was hoofdzakelijk te wijten aan het frequent voorkomen van vreemde of misleidende skills in de clusters. Zo signaleren de beoordelaars de gevoeligheid voor paroniemen (lexicaal sterk gelijkende termen) die inhoudelijk of functioneel niets met elkaar te maken hebben (e.g., “schrijfvaardig” en “afschrijvingen”). Daarnaast bleek de methode gevoelig voor de lengte van de skillbeschrijvingen. Dergelijke fouten vertaalden zich in lagere scores op clusterredundantie en -onderscheidbaarheid, en bijgevolg de globale kwaliteit.

Conclusie

Over het algemeen kon uit de beoordeling worden afgeleid dat methode C om minder intensieve menselijke interventie of methodologische verfijningen vraagt. Op basis van de feedback konden de volgende specifieke verbeterpunten geïdentificeerd worden:

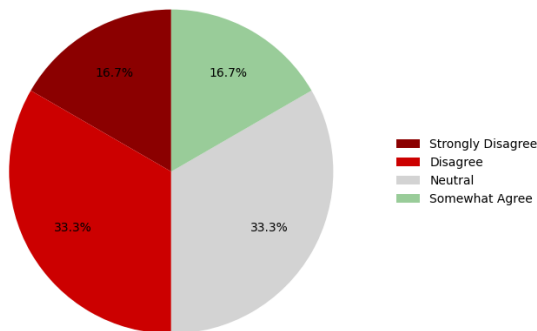
- **Vermindering van overgranulariteit (hoeveelheid clusters en skills).** Hoewel de data-gedreven aanpak poogde exhaustief te zijn en in volgende iteraties de clustering doorgevoerd kan worden op hoger hiërarchisch niveau, resulteerde de hoeveelheid weerhouden skills in een te groot aantal clusters. Het aantal clusters bleek het grootste struikelblok in de gebruiksvriendelijkheid en werd aangehaald als reden van onvermijdelijke overlap tussen clusters. Toekomstige inspanningen moeten verfijnd worden om tot een behapbaar aantal clusters te komen dat de verschillende facetten van het beroep overzichtelijk weergeeft. Gegeven dat sommige skills als irrelevant, perifeer of niche werden geacht, is het gebruik van een stringenter frequentiedrempel sterk aanbevolen. Van prioritair belang is het mappen van synoniemen om de redundantie te beperken en de clusterzuiverheid te bevorderen.
- **Corrigeren van semantische en linguïstische artefacten.** Methode C vertoonde specifieke technische fouten waarbij skills gegroepeerd werden op basis van tekstenmerken in plaats van inhoudelijke betekenis. De methode heeft baat bij het verhelpen van verkeerdelijke toekenningen voortkomend uit paroniemen (e.g., schrijfvaardig en afschrijvingen), infinitieven (rekenen, tekenen, zagen) en tekstlengte.
- **Verbetering in gebruiksvriendelijkheid op basis van labels.** De beoordelaars gaven herhaaldelijk aan dat clusterlabels aanzienlijk zouden helpen om de clusters te interpreteren en een overzicht te krijgen van de totale skills, al werd hier eveneens erkend dat dit kwalitatieve expertise zou kunnen vereisen.

- **Behandeling C1-competenties.** C1-competenties werden weerhouden om de vraag naar competenties zo volledig mogelijk te dekken, maar beïnvloedden volgens beoordelaars de kwaliteit van het werk. In volgende iteraties is het aanbevolen om de omzetting van de C1-competenties naar de AI-skilltags te evalueren. Dergelijke omzetting zou methode A en B ten goede kunnen komen.
- **Nieuwe versie Skill Cloud.** Het AI-team liet weten dat de volgende versie van Skill Cloud meer zou evolueren richting de C2-competenties en de skill-extractor beter rekening zou houden met de context van de vacatures. Dergelijke aanpassingen zouden kunnen helpen bij het verbeteren van de relevantie van de skills, het reduceren van synoniemen, en het vermijden van vreemd ogende skills (e.g., “verzoenen” kan verwijzen naar het reconciliëren van financiële gegevens).

Luik 2: Data-gedreven versus expert-gedreven taxonomie

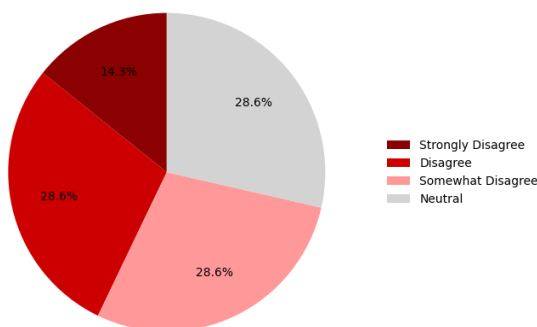
In dit luik werd gepeild naar de waarde van de data-gedreven taxonomie in verhouding tot de (competentiesets in) Competent 2.0-taxonomie. De items gingen na of de data-gedreven clustering nieuwe inzichten biedt, waardevol is als aanvulling op de competentiesets, wat de sterktes/zwaktes zijn t.o.v. Competent 2 en de bereidheid om dergelijke data-gedreven methoden aan te wenden in de toekomst. Er werd aan de beoordelaars gevraagd om in de vergelijking uit te gaan van de methode die zij als beste inschatten.

1. Nieuwe inzichten t.a.v. Competent 2.0



De data-gedreven methode bood nieuwe inzichten op het gebied van fijnmazigheid, voornamelijk door het identificeren van softwaregerelateerde skill tags die in de huidige taxonomie minder prominent aanwezig zijn. Ook het samenbrengen van inhoudelijk verwante thema's zoals loonverwerking of voorraadbeheer werd nuttig geacht voor aggregatie-doeleinden. Echter, fragmentatie en “ruis”, voortkomend uit skills die volgens beoordelaars niet bij het beroep horen, overschaduwden vaak de epistemische meerwaarde. Daarbij gaven beoordelaars ook aan over onvoldoende kennis te beschikken en dat het tijdsbestek te kort was om grondige uitspraken te kunnen doen. Het idee werd geopperd om nieuwe skills afzonderlijk aan te bieden.

2. Waardevol als aanvulling op de competentiesets in Competent 2.0



De scores op dit item zijn overwegend laag tot neutraal, wat wees op een beperkte overtuiging dat de clusters een waardevolle aanvulling zijn op de competentiesets. Voornamelijk het aantal skills en clusters, de wisselvallige samenhang, en het gebrek aan labels, ondersteunden deze antwoorden. De beoordelaars merkten op dat de

competentiesets een andere insteek hebben en het twee zaken zijn die naast elkaar bestaan. Er werd evenwel gesteld dat de data-gedreven aanpak zinvol kan zijn voor inhoudelijke aggregatie. Hoewel het aantal clusters als ongebruiksvriendelijk werd ervaren, gaf een beoordelaar aan dat de clusters desalniettemin nuttig kunnen zijn achter de schermen.

3. Sterktes en zwaktes t.o.v. de Competent 2.0-taxonomie

Sterktes	Zwaktes
Skilltags additionele informatie, vnl. software	Onvoldoende reductie van informatie
Bepaalde aggregaties zinvol (e.g., voorraadbeheer, loonverwerking)	Lagere interne samenhang
Korter geformuleerde skilltags	Niet essentiële skills (rijbewijzen, talen, bepaalde software)
Kan helpen bij het wegblijven van taak- of activiteitsgerichte beschrijvingen	Mix van C1, skills, knowledges, gedragsindicatoren
	Afwezigheid clusterlabels
	Lacunes (hergebruik vacatures, voornamelijk “nice-to-haves”, onderraportage “must-haves”, risico achterstalligheid beroepenclassificatiesysteem) zorgen ervoor dat vacaturedata niet als enige bron voor het opstellen van een taxonomie kan fungeren

4. Adoptiebereidheid

De bereidheid om de data-gedreven benadering in de toekomst in te zetten voor het onderhoud of de uitbreiding van de taxonomie (VDAB) of als basis voor het geïntegreerd arbeidsmarktprojectiemodel op geaggregeerd competentieniveau (Steunpunt Werk) is voorwaardelijk positief. Hoewel de meeste beoordelaars potentieel zien, wordt de huidige output als te "ruw" beschouwd voor onmiddellijke implementatie. Vermits de hierboven geïdentificeerde kernproblemen verholpen kunnen worden, zien beoordelaars vooral meerwaarde als ondersteunend instrument voor het detecteren van nieuwe of ontbrekende skills en als inspiratiebron voor alternatieve (kortere) competentiebeschrijvingen. Met het oog op de beroepenprojecties geniet voor het Steunpunt Werk een niet-beroepsgebonden clustering de voorkeur, aangezien dit de nodige aggregaties en onderlinge vergelijkingen tussen beroepsgroepen faciliteert.

Conclusie

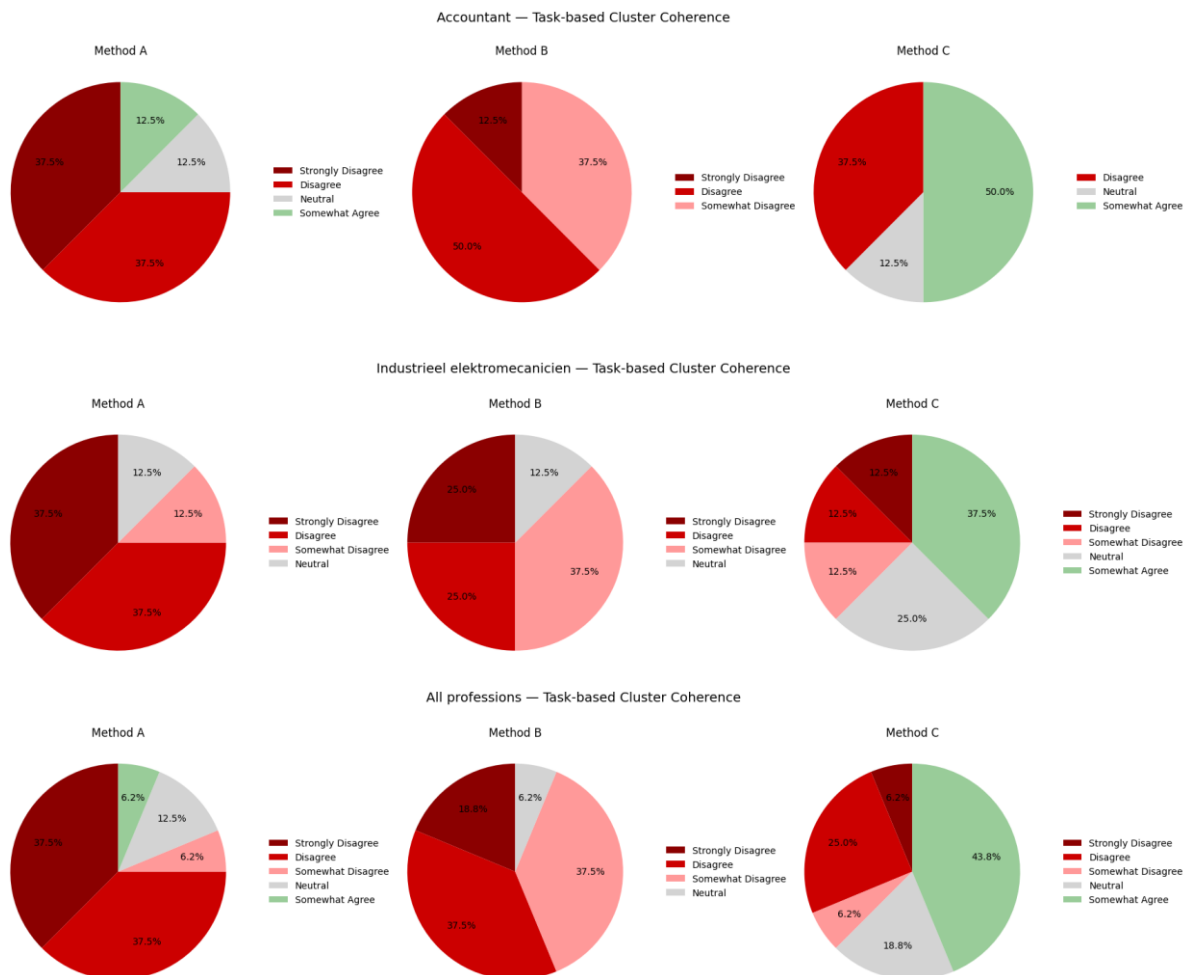
De resultaten van deze bevraging tonen ondubbelzinnig aan dat een volledig autonome, data-gedreven clusteringaanpak momenteel onvoldoende kwaliteit biedt om zelfstandig als competentietaxonomie te fungeren. Hoewel de semantische benadering (Methode C) een duidelijke vooruitgang markeert in inhoudelijke herkenbaarheid, blijft deze methode structureel tekortschieten op het vlak van taakgerichte coherentie, interpretatie en algemene bruikbaarheid. De verzamelde feedback biedt echter concrete handvatten voor toekomstige iteraties, waarbij verbeterpunten zoals het reduceren van de overgranulariteit en synoniemen, het toekennen van labels en het corrigeren van linguïstische artefacten centraal staan.

De bevindingen bevestigen dat de inzet van menselijke experts en gerichte kwalitatieve interventies noodzakelijk en onvervangbaar blijven om tot betekenisvolle, inzetbare structuren te komen. Bijgevolg wordt de data-gedreven benadering niet gepositioneerd als een vervanging van het huidige instrumentarium voor taxonomie-ontwikkeling, maar veeleer erkend als een waardevol ondersteunend instrument dat detailinformatie en inspiratie kan bieden bij de detectie van nieuwe trends en de verfijning van competentiebeschrijvingen.

Appendix

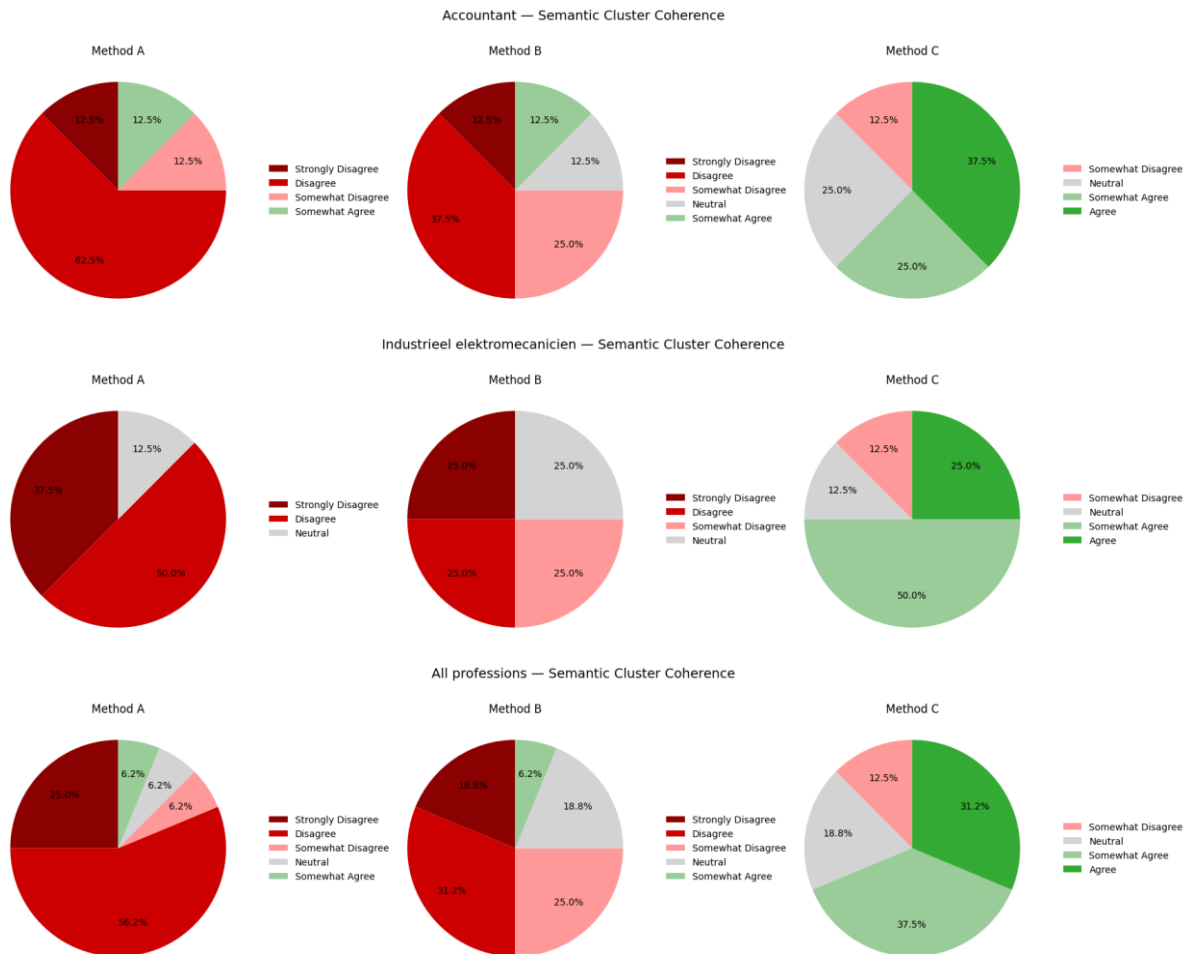
Taakgerelateerde clustercoherentie

Stelling: De skills in de clusters zijn duidelijk gezamenlijk vereist om een bepaald takenpakket succesvol te kunnen uitvoeren.



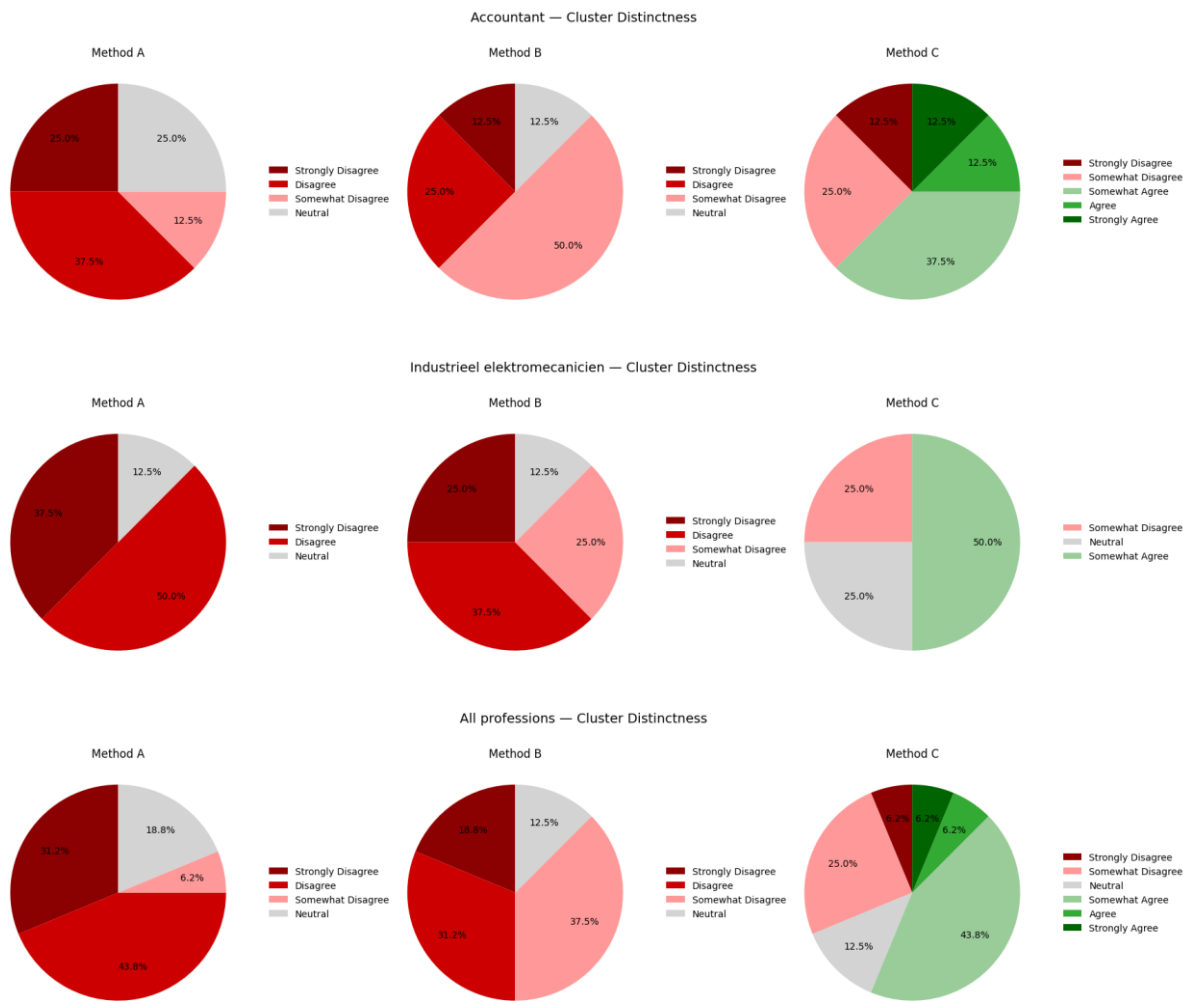
Semantische clustercoherentie

Stelling: De clusters bevatten inhoudelijk/semantisch sterk verwante skills die logisch bij elkaar passen.



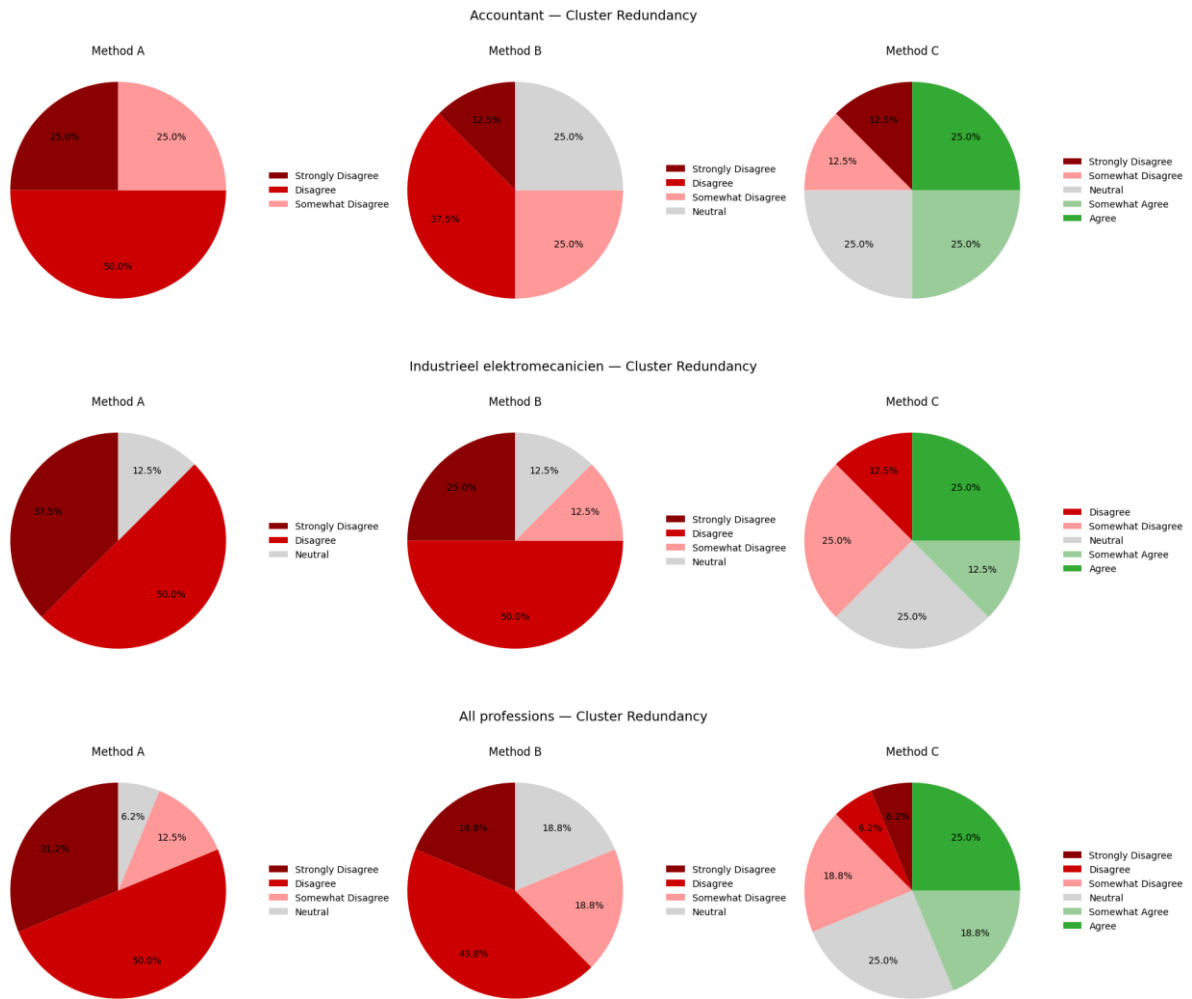
Cluster onderscheidbaarheid

Stelling: De clusters zijn voldoende van elkaar te onderscheiden qua inhoud (type skills) of functie (rol in taken of werkprocessen).



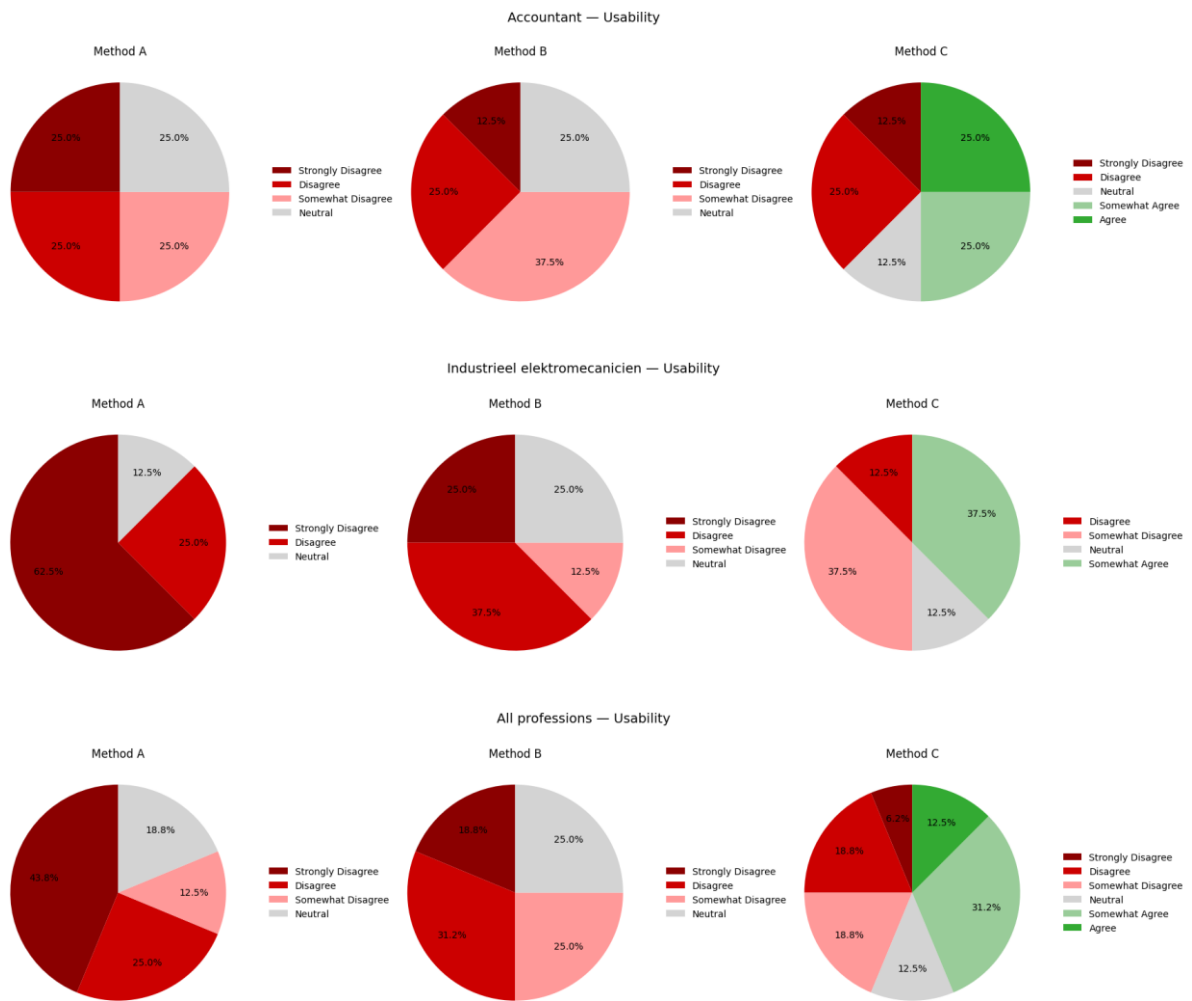
Cluster redundantie

Stelling: De clusters bevatten zelden skills die ongepast of niet logisch lijken binnen hun groep.



Bruikbaarheid clusters

Stelling: De clustering helpt mij om een duidelijk overzicht te krijgen van de verschillende soorten vaardigheden binnen dit beroep.



Algemene kwaliteit

Stelling: Over het geheel genomen vind ik de clustering van hoge kwaliteit voor dit beroep.

