

**De antwoorden op de schriftelijke vragen die via de website zijn gesteld zijn te vinden op de [discussiepagina](#).**

## **Meetup 30 juni**

### **Hoe kunnen we geautomatiseerd en doorlopend de 690 cao's in Nederland analyseren?**

*Zit het grootste probleem in het verwerken van een nieuwe versie van een bestaande cao's of het verwerken van nieuwe cao's?*

Het ene probleem is niet groter dan het ander: beiden zijn belangrijk. Bestaande cao's, dus cao's die al eens door ons zijn onderzocht, zijn al een keer verwerkt in ons systeem, dat is het verschil met nieuwe cao's. Vaak veranderen cao's gedeeltelijk wanneer er nieuwe afspraken worden gemaakt tussen werkgevers en werknemers. In die zin is 'nieuw' een relatief begrip. Het verschil met de helemaal nieuwe cao's is dat cao de nieuwe cao's dus in zijn geheel nog niet gecodeerd en geanalyseerd zijn. De nieuwe cao's zijn nu niet gecodeerd en geanalyseerd omdat we nu niet genoeg tijd hebben om alle cao's door te nemen.

*Kun je misschien een paar voorbeelden geven van inzichten/datapunten die jullie uit de cao's halen?*

- Voorbeeld 1: We kijken in de cao's of er afspraken in zijn gemaakt over een individuele ontwikkelbudgetten en hoe de bijbehorende budgetten zich ontwikkelen. Dit doen we omdat beleidsmakers met de sociale partners (de werkgevers- en werknemersorganisaties) hebben afgesproken dat dit via cao's gestimuleerd wordt. Voorafgaand aan de beleidsafspraken met de sociale partners hebben we ook een nulmeting op dit onderwerp gedaan.
- Voorbeeld 2: In de afgelopen tijd is door beleidsmedewerkers gekeken naar de mogelijkheid om een minimum uurloon in te voeren. Met onze cao analyse brengen we in kaart in welke sectoren een wettelijk minimum uurloon mogelijk zou wringen, omdat de huidige uurlonen hier bijvoorbeeld ver onder liggen, vanwege bijvoorbeeld vergoedingen van tijd waarin werknemers opgeroepen kunnen worden.

*"Wij moeten kunnen uitleggen waarom een bepaalde waarde of score aan een cao-(deel)tekst is toegekend. Zodat wij altijd kunnen toelichten hoe wij tot onze conclusies komen." --> op welk niveau van detail moet dit door de techniek worden uitgelegd? Is het voldoende dat de techniek verwijst naar de alinea waar de oorspronkelijke tekst gevonden kan worden?*

Voor de onderzoekers bij SZW moet te achterhalen zijn waarom een waarde is toegekend, zodat dit ook eventueel handmatig gewijzigd kan worden. Kunnen verwijzen naar de alinea waar de oorspronkelijke tekst staat helpt daarbij, maar we willen ook begrijpen waarom het systeem daar vervolgens een bepaalde waarde aan toekent.

*Moet er gewerkt worden binnen de bestaande kaders van de Wet cao en / of de Wet AVV of bestaan er mogelijkheden om ook buiten die kaders te werken?*

De cao's vallen allemaal onder de wet cao en soms zijn zij algemeen verbindend verklaard, dan is ook de wet AVV van toepassing. Op de analyse van de cao teksten zelf heeft dit verder geen invloed. Voor het automatiseren van de analyse maken deze wettelijke kaders dus geen verschil.

*Het handmatig verwerken dat jullie nu doen, hoe ziet dat eruit? Wat en hoe doen jullie dat nu?*

We werken met een programma (Max-QDA) waarmee we aan stukken tekst uit de cao codes kunnen hangen. Dit doen we grotendeels handmatig. Soms is het mogelijk om een zoekwoord te gebruiken en hier automatisch een code aan te hangen maar over het algemeen is de zoekfunctie van het programma (Max-QDA) niet toereikend voor het automatisch toekennen van codes. Na het toekennen van de codes aan de teksten administreren we in Excel of SPSS wat de waarde is van deze code. Nadat alle cao's zijn gecodeerd en de waarden van de codes geadmineerd doen we analyses op het Excel of SPSS bestand. Daar maken we vervolgens een dashboard van en een rapport.

*In de uitvraag wordt gerefereerd aan "codes" die de betekenissen van de cao's dekken. Kunnen jullie daar een voorbeeld van geven?*

Een voorbeeld is het individuele leer- en ontwikkelbudget: als hierover een afspraak is opgenomen in een cao hangen we daar een code aan. Vervolgens kijken we ook meer in detail, bijvoorbeeld naar de hoogte van het bijbehorende budget of waar dit wel en niet aan besteed kan worden. Zie voor een nadere toelichting het antwoord op de volgende vraag.

*Worden dezelfde codes voor alle cao's gebruikt? Is er een voorbeeld lijst van deze codes?*

We hebben losse lijsten per onderzoek, omdat elk onderzoek dat we doen gericht is op een specifiek onderwerp (zoals flexibele contracten, leren en ontwikkelen, etc.). Hier hebben we voorbeelden van, die kunnen we delen. Met de challenge willen we één onderwerp als voorbeeld nemen: interne / externe flexibiliteit. We beginnen een project met het maken van een vragenlijst in overleg met de beleidsmedewerkers. Vervolgens worden hierop de codes in Max-QDA gebaseerd. Deze worden omgezet naar een Excel sheet en geanalyseerd in SPSS of gevisualiseerd in Power BI. De vragenlijst en opgeschoonde Excel sheet (genaamd Fact sheet) hadden we de vorige keer bijgevoegd bij het projectplan. Als nieuwe bijlage hebben we de lijst met codes uit Max-QDA geëxporteerd met informatie over aantallen. Deze is aan het einde van dit document te vinden.

De Excelbestanden hebben een hoger detailniveau dan de Max-QDA-bestanden, zie ook de voorbeeldbestanden. De codes van de Max-QDA-bestanden verwijzen naar een alinea in de cao-tekst. In het Excelbestand noteren we bijvoorbeeld: lonen, de hoogte van bedragen en waar ze aan uitgegeven kunnen worden. In deze challenge willen we met het automatiseren van de tekstanalyse idealiter het detailniveau van de Excel bestanden bereiken. Maar ook het automatisch zoeken en coderen van alinea's waar de informatie gevonden kan worden zou al tijdswinst opleveren.

*Zijn jullie op zoek naar het beantwoorden van specifieke vragen die jullie hebben over de cao's (bijvoorbeeld leerbudgetten) of zijn jullie ook geïnteresseerd in welke thema's, die wellicht nog niet in beeld zijn, behandeld worden in cao's?*

De specifieke vragen die wij hebben over de cao's zijn leidend. Thema's die nog niet in beeld hebben halen we vrijwel niet uit cao's. We halen dit uit gesprekken met beleidsmedewerker en uit de principe akkoorden, die door de sociale partners zijn gesloten. Wel zou het een aanvullende kunnen zijn, vanuit waar nieuwe thema's omhoogkomen.

*Is het de bedoeling dat de data ook publiekelijk ter beschikking wordt gesteld i.v.m. het Europeesrechtelijk kenbaarheidsvereiste van (harde kern) arbeidsvoorwaarden?*

Ja: We publiceren momenteel al onze data op het open data portal van de rijksoverheid. Dit moet mogelijk blijven. Dit geldt in elk geval voor de gekwantificeerde data. Voor de toekomst zou het een mooie ontwikkeling zijn om meer data openbaar te kunnen maken. Cao zelf zijn al openbaar.

*Is het dan bijvoorbeeld voldoende dat de techniek wijst naar een alinea binnen de cao waar het antwoord gevonden kan worden?*

Zie het antwoord op de eerdere vraag hierover.

*Ik zag dat jullie een voorkeur hebben voor een webapplicatie. Hebben jullie budget voor hosting, en updates, en hebben jullie een voorkeur voor bepaalde platforms/clouds (Windows, Linux, Amazon Cloud, etc.) of kan dit overal?*

Onze voorkeur gaat uit naar een webapplicatie omdat het draaien van een applicatie op onze eigen servers heel veel tijd zou kosten. Daarvoor is namelijk heel veel overleg en testen nodig. Dit zou zonde zijn voor het tempo van ontwikkeling van de oplossing. We stellen de volgende vereisten aan een webapplicatie:

- De data/gegevens moeten in Nederlad staan en een verwerkerovereenkomst is verplicht.
- In de toekomst worden beveiligingsaudits verplicht en moet worden voldaan aan Baseline Informatiebeveiliging overheid (BIO) – dit is de standaard beveiligingsrichtlijn van de overheid.
- De applicatie moet de laatste 2 versies van Edge, FireFox ondersteunen.
- Het ter beschikking gestelde budget is inclusief hosting, beheer, licenties, eenmalige projectkosten en doorlopend onderhoud.

- We gaan er vanuit dat de applicatie in elk geval gedurende de looptijd van de pilot blijft werken/voldoende lang blijft lopen om te testen of de applicatie geschikt is.

*Is men bereid om extra eisen aan de sociale partners te stellen i.v.m. het opmaken en aanleveren van cao's?*

Dit is niet mogelijk: het is geheel aan de sociale partners (de werkgevers- en werknemersorganisaties) hoe zij de afspraken vastleggen. Vereisten wat betreft opmaak en aanlevering kunnen wij niet opleggen. We hebben in het verleden wel themadagen georganiseerd waarin we een adviserende en faciliterende rol hebben gespeeld op het gebied van opmaak en aanlevering, leesbaarheid etc. van cao's.

*Hoeveel tijd zijn jullie nu kwijt aan het analyseren van één cao? Hoeveel sneller willen jullie dat dit gaat na deze opdracht?*

Momenteel zijn wij per onderzoek 1,5 uur bezig met het doorlichten van één cao (coderen in Max-QDA en administreren van de waarden in Excel of SPSS). Omdat wij ongeveer 20 onderzoeken hebben betekent dit dat we per cao 30 uur nodig hebben. Deze tijd willen we graag flink terugbrengen door het coderen en administreren te automatiseren. Daarbij is niet alleen snelheid maar ook de kwaliteit belangrijk. Want als we veel moeten controleren en corrigeren dan is de tijdswinst beperkt.

Per onderzoek zijn we nu bijna 4 weken kwijt aan het coderen en administreren. En vervolgens twee weken aan het schrijven van het rapport. We willen deze verhouding omdraaien: we willen toe naar een werkwijze waarbij we meer tijd besteden aan analyse en rapportage en minder tijd aan het coderen van cao's.

*Zouden jullie in de toekomst ook automatisering willen zien in de analyse van alle relevante informatie?*

We willen nu beginnen (in het Start-up in resident-traject) met één onderzoek/onderwerp: interne en externe flexibele contracten. Als dit goed werkt willen we dit verder uitrollen naar alle onderwerpen waarop wij onderzoek doen.

*Er is vaste set van coderingen binnen alle cao's die in kaart gebracht wordt; dit is handwerk (klopt dit?); hoe gaan jullie om met het (terugwerkende kracht) handwerk van nieuw toegevoegde coderingen?*

Er worden inderdaad een vaste set van codes gebruikt per onderwerp waarop wij onderzoek verrichten. Het codereen gebeurt nu inderdaad handmatig. Indien er een nieuwe code wordt toegevoegd doen we dit doorgaans niet met terugwerkende kracht. Dat is namelijk te veel werk en vaak ook niet interessant. De code wordt namelijk toegevoegd omdat dat (deel)onderwerp interessant is geworden. Dat betekent doorgaans dat het onderwerp voor die tijd niet genoemd werd in de cao.

*Welk beeld hebben jullie bij de toekomstige wensen/ideeën die nu nog in de kinderschoenen staan zoals in de presentatie werd genoemd?*

We willen uiteindelijk toe naar een interactievere presentatie van de data, zodat niet alleen een rapport beschikbaar is maar een interactief dashboard. Dit valt momenteel buiten de scope van de opdracht.

*Kunnen jullie voorbeeld materiaal delen [cao's die reeds gecodeerd zijn]? Zodat deze ook gebruikt kunnen worden om de modellen te trainen?*

Dit materiaal is beschikbaar en kan tijdens de Start-up in-residence periode gebruikt worden.

*Hoeveel cao's plus koppelingen zijn dat nu?*

Momenteel analyseren we 100 cao's, waarbij we in het Start-up in Residence traject willen inzoomen op één onderwerp: interne en externe flexibele contracten. Op dit onderwerp gebruiken we momenteel gemiddeld 10 codes per cao. In totaal komt dit voor het hele onderwerp neer op ca. 1000 segmenten. Zie bijlage.

*Wat betreft het co-creatie proces. Welke rol zien jullie daarin voor jullie zelf? Welke mate van betrokkenheid vinden jullie wenselijk?*

We willen graag meegenomen worden in het proces dat de geselecteerde start-up gaat doorlopen. Op die manier wordt de werking van de oplossing voor ons inzichtelijk: dat vinden we belangrijk. Het gaat dus echt om een gelijkwaardige samenwerking.

*Wat is de definitie van een start-up?*

Er wordt geen definitie van een start-up gehanteerd. We zoeken vooral jonge bedrijven in Nederland die voldoende affiniteit hebben met de challenge. De passendheid van de oplossing / kwaliteit van de ideeën is leidend in het selectieproces. Wel dient de Start-up geregistreerd te zijn bij de KvK.

*Wat voor organisatie hebben jullie zelf de voorkeur om mee te werken?*

Wij hebben geen voorkeur voor een specifiek formaat of mate van senioriteit van het bedrijf (bijvoorbeeld MKB of start-up): het gaat echt om de kwaliteit van de oplossing.

*Waar sturen we eventuele follow-up vragen naar toe?*

Na deze meetup kunnen schriftelijke vragen gesteld worden tot 6 juli, 12:00 uur. De antwoorden op de vragen die tijdens deze meetup worden gesteld, zullen na deze deadline worden gepubliceerd, samen met de antwoorden op de vragen die schriftelijk worden gesteld. De schriftelijke vragen kunnen gesteld worden via Starthubs.

Worden alle vragen en antwoorden en de presentatie op de challenge pagina gepost?

*Ja*