

Hoe GenAI de informatie-extractie uit complexe industriële documenten verbetert

29 juni 2026, 12:00

Sreeraj Rajendran

Nicolás González-Deleito

Ali Beigrezaei

Aziz Nebli

Van technische documenten naar bruikbare informatie

Industriële processen, fabrieksindelingen, installatieconfiguraties of onderdelenspecificaties worden doorgaans vastgelegd in een brede waaier van technische bronnen: tekstuele beschrijvingen, schema's, figuren, tabellen en afbeeldingen. De relevante informatie uit deze documenten extraheren is essentieel voor activiteiten zoals onderhoud, audits, certificeringen en de creatie van digitale tweelingen. Het blijft echter een complexe taak om gegevens uit dergelijke heterogene bronnen op betrouwbare wijze te identificeren, met elkaar in verband te brengen en te interpreteren.

Via het internationale [NARRATE-project](#), ondersteund door het Brussels Hoofdstedelijk Gewest en Innoviris, onderzoeken Sirris en 3E hoe generatieve AI deze informatie automatisch kan extraheren, koppelen en structureren – en hoe het de weg vrijmaakt voor snellere en efficiëntere verwerkings- en automatiseringsworkflows.

CAD drawing

Over NARRATE

[NARRATE](#) is een internationaal onderzoeksproject en verenigt veertien partners uit vijf landen: België, Estland, Zuid-Korea, Spanje en Turkije. Het project heeft als doel betrouwbare en ethische, door AI aangestuurde interfaces voor menselijke interactie te ontwikkelen, waarbij multimodale gegevensintegratie en geautomatiseerde kennisupdates

hand in hand gaan.

[Lees meer](#)

Waarom informatie extraheren nog steeds moeilijk is

De complexiteit schuilt niet zozeer in het vinden van afzonderlijke stukjes informatie in verschillende bronnen, dan wel in het begrijpen hoe deze allemaal met elkaar verband houden:

- De input is vaak ongestructureerd en aangeboden in verschillende formaten (bv. tekst, tabellen, afbeeldingen)
- De informatie is via verschillende formaten met elkaar verbonden (bv. een tekstuele beschrijving die verwijst naar informatie in een tabel of afbeelding)
- Referenties staan vaak ver uit elkaar binnen hetzelfde document of zijn verspreid over meerdere bestanden
- Hetzelfde concept kan worden beschreven met behulp van verschillende naamgevingsconventies of identificatiecodes

Daardoor is informatie verzamelen slechts een deel van de moeilijkheid. Systemen moeten bovendien verbanden leggen, informatie samenvoegen en de context behouden binnen heterogene documentverzamelingen.

Heterogene documenten omzetten in bruikbare gegevens

In het kader van [NARRATE](#) ontwikkelen Sirris en 3E AI-gestuurde pijplijnen die automatisch informatie uit diverse technische documenten kunnen extraheren, koppelen en structureren. Daartoe combineren Sirris en 3E twee complementaire AI-strategieën.

1. Lokale LLM's

Sirris richt zich op lokaal gehoste taalmodellen (LLM's) zoals Llama, Gemma en Mistral. Door modellen op de interne infrastructuur te hosten, garandeert men dat gevoelige klantgegevens binnen de bedrijfsomgeving blijven. De eerste workflow hanteert een meerstappenbenadering.

1. Bepaal het documenttype en de inhoud.
2. Extraheer de relevante informatie.
3. Structureer de geëxtraheerde gegevens.
4. Herhaal de bovenstaande stappen voor elk document.
5. Voeg de informatie uit de verschillende bestanden samen tot één enkele output.

Naarmate het project vorderde, bleek het koppelen van informatie uit verschillende documenten de grootste uitdaging te zijn. Dat leidde tot een verschuiving naar een op agents gebaseerde

workflow, waarbij AI de verschillende stappen coördineert die nodig zijn om informatie binnen een volledige documentset te detecteren, te koppelen en te valideren.

Dankzij deze agentgebaseerde workflow verwerkt het systeem probleemloos veelzijdige, multimodale documentsets met een combinatie van tekst, tabellen, afbeeldingen en schema's in uiteenlopende formaten en met verschillende naamgevingsconventies.

2. Cloudondersteunde LLM's

Tegelijkertijd ontwikkelt 3E een cloudgebaseerde pijplijn binnen een beveiligde Azure-omgeving. De documenten worden eerst verwerkt en op intelligente wijze opgedeeld met behulp van technologieën zoals Azure Computer Vision en Document Intelligence. GenAI-modellen identificeren vervolgens de informatietypes en proberen de onderlinge relaties automatisch te reconstrueren. Het doel is hetzelfde: complexe technische documentatie omzetten in een gestructureerd, gebundeld overzicht, met zo weinig mogelijk handmatige tussenkomsten.

The cloud

De betrouwbaarheid verbeteren door middel van validatie

Om de betrouwbaarheid te verbeteren, hebben beide partners uitgebreide validatiemechanismen uitgewerkt. Sirris gebruikt Langfuse om prompts te monitoren, modelparameters te vergelijken en de kwaliteit van de output te evalueren. 3E hanteert een ensemble-benadering waarbij meerdere LLM's dezelfde extractietaak parallel uitvoeren. De verkregen resultaten worden vervolgens gecontroleerd op inconsistenties en ontbrekende of dubbele informatie, alvorens ze definitief worden samengevoegd.

Hoe meet u succes?

Tot slot bleef er nog één belangrijke vraag over: hoe beoordeelt men de kwaliteit van automatisch geëxtraheerde informatie? De twee partners hebben op maat gemaakte evaluatiemaatstaven ontwikkeld die gericht zijn op:

- Volledigheid
- Aanwezigheid van de juiste items
- Kwaliteit in het veld
- Numerieke juistheid

De eerste resultaten zijn veelbelovend. De informatie-extractie levert nu al uitstekende resultaten op. Automatisch complexe verbanden leggen tussen informatie-items blijft echter het moeilijkste aspect.

Wat zijn de volgende stappen?

Uit de eerste fase van het project is al gebleken dat GenAI het extraheren van technische informatie uit complexe technische documentatie aanzienlijk kan versnellen. De volgende stap bestaat erin deze mogelijkheden verder uit te breiden door complexe koppelingen tussen componenten mogelijk te maken en de verkregen kennis om te zetten in praktische ondersteuning bij besluitvorming. Het consortium zal zich ook toeleggen op toepassingen zoals intelligente dashboards, aanbevelingssystemen en geautomatiseerd kennisbeheer ter ondersteuning van operationele besluitvorming.

De reis is nog lang niet ten einde. Toch laat de vooruitgang die binnen [NARRATE](#) is geboekt, nu al zien hoe complexe technische documentatie kan worden omgezet in bruikbare digitale informatie. Naarmate GenAI zich verder ontwikkelt, zou betrouwbare informatie-extractie een belangrijke factor kunnen worden om processen te verbeteren, met onder meer een snellere onboarding, grotere schaalbaarheid en efficiënter activabeheer als resultaat.

Benieuwd hoe GenAI uw eigen technische workflows zou kunnen stroomlijnen?

Sirris helpt bedrijven om AI en GenAI toe te passen op complexe industriële uitdagingen, variërend van informatie-extractie en kennisbeheer tot automatisering en beslissingsondersteuning.

[Contacteer ons](#)

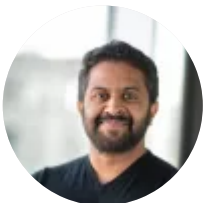
Partners

met de steun van ITEA4, Innoviris en het Brussels Hoofdstedelijk Gewest



Medegefinancierd door

Auteurs



Sreeraj Rajendran



Nicolás González-Deleito



Ali Beigrezaei



Aziz Nebli