



2^e Nota van Inlichtingen SBIR Historische transcripties met AI voor Nationaal Archief 2019

Versie 18 september 2019

Deze nota bevat de vragen die na de informatiebijeenkomst zijn gesteld. Deze zijn via e-mail zijn binnengekomen met de antwoorden die daarop zijn gegeven of naderhand geformuleerd zijn.

	Vraag	Antwoord
1.	Valt het verbeteren van het transcribeer proces binnen de scope van het SBIR project?	Het transcriberen en uitvoeren van automatische handschriftherkenning valt buiten scope van deze SBIR.
2.	Voor het toepassen van AI is trainingsdata nodig. Het transcriptieplatform Transkribus gebruikt voor haar Machine Learning model ook getrainde data samengesteld door projectmedewerkers. Heeft het Nationaal Archief ook nagedacht wie Trainingsdata kan verzorgen voor het gevraagde prototype en later in de beheer fase? Zijn daarvoor Nationaal Archief medewerkers beschikbaar of moet dat anders ingevuld worden?	Het NA heeft een transcriptieorganisatie ingericht voor het vervaardigen van transcripties voor de Machine Learning. Het vervaardigen van transcripties en trainingsmodellen is buiten scope van de SBIR.
3.	Als er op basis van de beschikbaar gestelde Transcriptie in het vervolgproces (dat binnen de scope valt) ook machine learning plaatsvindt om bijvoorbeeld Named Entity Recognition (NER) uit te voeren, dan is daarvoor ook trainingsdata nodig. Wie is daarvoor beschikbaar?	Het Nationaal Archief heeft geen personeel beschikbaar voor het maken van trainingsdata voor bijvoorbeeld NER voor het gevraagde prototype. Het Nationaal Archief heeft voor het vervaardigen van trainingsdata voor de automatische handschriftherkenning een team opgezet, zij maken transcripties. Hun taken houden wat het Nationaal Archief op zodra het gewenste aantal transcripties gemaakt is.
4.	Wordt de Transcriptie Beheersysteem beoordeeld/getest?	We kijken naar hoe het beheren van transcripties wordt uitgewerkt. Een compleet transcriptiebeheersysteem en het testen daarvan is pas later aan de orde in fase 3.



	Vraag	Antwoord
5.	Wat zijn de onderdelen waarop het Prototype (fase 2) wordt beoordeelt? Bv: gebruiksvriendelijkheid, integratie met andere systeemonderdelen, herbruikbaarheid voor fase3, extra features, aansluiting bij moscow lijst etc.	Aan het eind van fase 2 wordt het prototype beoordeeld aan de hand van de drie criteria en de toelichting daarbij uit de handleiding en de oproep.
6.	Hoe wordt het Prototype (objectief) getoetst?	Aan het eind van fase 2 laat u een demonstratie zien van het werkende prototype. Aan de hand van deze demonstratie en het eindrapport wordt getoetst of u de opdracht in voldoende mate heeft uitgevoerd.
7.	Wordt de API beoordeeld/getest?	Nee, deze zal pas worden getest in fase 3.
8.	Wordt de performance beoordeeld/getest? Zijn hiervoor criteria? Snelheid en/of kwaliteit zoekresultaten?	Nee
9.	Moeten we ook modules opleveren aan Opdrachtgever, zodat deze door de Opdrachtgever zelf geïnstalleerd kunnen worden	Nee
10.	Wordt de documentatie (ontwerp, handleidingen) beoordeeld/getest?	Nee
11.	Gaat Opdrachtgever ook tijdens het project testen (en feedback geven)?	Nee
12.	Welke talen moeten ondersteund worden in de UI prototype?	(oud) Nederlands
13.	Welke talen komen voor in de data?	De transcripties zijn bijna allemaal in het Nederlands. Af en toe zal er een document in het Frans of een andere taal aanwezig zijn.
14.	Welke talen moeten ondersteund worden in fase2	Het prototype moet functioneren met Nederlands. Het product van fase 3 moet functioneren met Nederlands en Engels, zie website Nationaal Archief. Er wordt echter niet verwacht dat hiervoor inspanningen moeten worden verricht door Leverancier.
15.	Is het gebruik van Jira en Confluence in fase 2 relevant?	Nee, helemaal niet relevant



	Vraag	Antwoord
16.	Wat beoogde het Nationaal Archief met de presentatie van de user journey?	Het Nationaal Archief vindt het belangrijk dat bij het maken van een functionaliteit voor zoeken in en tonen van transcripties er wordt gedacht vanuit de gebruiker. "Wat heeft de gebruiker nodig en hoe kan gericht zoeken de gebruiker helpen?"
17.	Bij vraag 80 van de eerste Nota van Inlichtingen antwoordt u dat de website binnen scope valt van het prototype. Mag dat ook een webachtige omgeving zijn?	Ja, voor het prototype hoeft geen sprake te zijn van een volwaardige website.
18.	En de eerste Nota van inlichtingen schrijft u dat de ground truth niet beschikbaar wordt gesteld omdat het verwaardigen van transcripties niet in scope is van deze SBIR. We verwachten echter dat ook voor machine learning bij bijvoorbeeld NER de transcripties van pas komen.	De transcripties en scans zullen beschikbaar worden gesteld in een derde datapakket.
19.	Hoe verloopt de communicatie tijdens fase2? (I.v.m. testen en agile/scrum werken)	zie vragen 20 en 21
20.	Worden de vragen/antwoorden tijdens het project gedeeld?	Het Nationaal Archief stelt voor dat er maximaal 1 x per 3 weken een overleg is. Dit zijn vaste vergadermomenten op locatie van het Nationaal Archief. alle leveranciers die een prototype maken zijn hiervoor uitgenodigd. Vragen van leveranciers worden een week van te voren ingediend. Het Nationaal Archief zorgt voor het schriftelijk beantwoorden van de vragen aan alle leveranciers die een prototype maken.
21.	Is het mogelijk om bijvoorbeeld tweewekelijks een demo aan de stakeholders te geven en feedback te vragen?	Het Nationaal Archief stelt voor drie maal de gelegenheid te geven een demo te geven van het concept van het prototype (medio november, december en januari). Andere leveranciers worden hierbij niet uitgenodigd. Vragen van leveranciers worden een week van te voren ingediend. Als er naar aanleiding van het tonen van de demo van het concept van het prototype informatie wordt versterkt door het Nationaal Archief die ook waardevol is voor de andere leveranciers die een



	Vraag	Antwoord
		prototype maken, dan zal deze informatie worden gedeeld aan alle leveranciers die een prototype maken.
22.	Worden de namen van deelnemers bekend gemaakt?	Nee, alleen de namen van de winnaars (de indieners die een opdracht krijgen) worden bekend gemaakt.
23.	Hoe lang duurt fase 3 en wanneer wordt deze aangekondigd?	Fase 3 gaat ongeveer 6 maanden duren en de aanbesteding start in februari/maart 2020.