**Student**

Daan Romijn

**Studentnummer****Organisatie**

Universiteit Utrecht

**Bedrijfsbegeleider****Datum**

27/10/2025

**Versie**

1.0

# AI voor topics binnen Ricgraph

## Afstudeervoorstel

## Inhoudsopgave

1.	<b>INLEIDING</b> .....	4
2.	<b>ORGANISATORISCHE CONTEXT</b> .....	4
3.	<b>KWESTIE</b> .....	4
4.	<b>DOEL</b> .....	5
5.	<b>RANDVOORWAARDEN</b> .....	5
6.	<b>AFBAKENING</b> .....	6
7.	<b>PRODUCTEN</b> .....	6
8.	<b>ICT-COMPONENT</b> .....	6
9.	<b>BEROEPSTAKEN</b> .....	6
	Beroepstaak 1 .....	6
	Beroepstaak 2 .....	7
10.	<b>AANTONEN ONDERZOEKEND VERMOGEN</b> .....	8
11.	<b>ETHISCHE AFWEGINGEN</b> .....	8
12.	<b>NIVEAU BEDRIJFSBEGELEIDER</b> .....	8
13.	<b>BIJLAGE</b> .....	9

## Algemene informatie

### OPDRACHT

Soort stage	<input type="checkbox"/> Stage (derdejaars) <input checked="" type="checkbox"/> Afstuderen		
Studierichting	<input type="checkbox"/> Artificial Intelligence <input checked="" type="checkbox"/> Business IT & Management <input type="checkbox"/> Cyber Security & Cloud		<input type="checkbox"/> Software Development <input type="checkbox"/> Technische Informatica
Variant studierichting	<input type="checkbox"/> Deeltijd <input type="checkbox"/> Duaal	<input type="checkbox"/> Datamanagement & BI (alleen BIM deeltijd) <input checked="" type="checkbox"/> Voltijd	
Afstuderen met een	<input type="checkbox"/> Beroepsproduct <input checked="" type="checkbox"/> Scriptie		
Waar heb je stagegelopen in je derde jaar? (voor afstudeerders)			

### STUDENT

Achternaam	Romijn	Voorletters	D
Studentnummer		Voornaam	Daan
E-mail		Telefoon	
Woonplaats			
Is de opdracht vertrouwelijk?	<input type="checkbox"/> Ja <input checked="" type="checkbox"/> Nee		
Is er sprake van een duo-stage?	<input type="checkbox"/> Ja <input checked="" type="checkbox"/> Nee		
SLB'er			

### ORGANISATIE

Organisatienaam	Universiteit Utrecht		
Postadres	Postbus 80125		
Postcode	3508 TC	Plaats	Utrecht
Land	Nederland		
Bezoekadres	Bestuursgebouw, Heidelberglaan 8		
Postcode	3584 CS	Plaats	Utrecht
Land	Nederland		
Telefoon		E-mail	
Aantal werknemers	7.282 (2024)	Website	uu.nl

### BEDRIJFSBEGELEIDER

Achternaam		Voorletters		<input type="checkbox"/> Mevr. <input type="checkbox"/> Dhr.
Afdeling	Directie Information and Technology Services			
Functie				
Werkzaam op HBO+ niveau?	<input type="checkbox"/> Ja <input type="checkbox"/> Nee			
LinkedIn-profiel				
Telefoon		Mobiel		
E-mail				

## 1. Inleiding

De Universiteit Utrecht laat zich in haar missie en visie leiden door kernwaarden als ambitie, betrokkenheid, inspiratie en onafhankelijkheid. Binnen deze context ondersteunt de universiteit onderzoekers, beleidsmakers en ondersteunend personeel bij het verkrijgen van inzicht in onderzoeksinformatie. Ricgraph is een systeem dat data uit verschillende bronnen samenbrengt en verrijkt met nieuwe verbanden.

Deze afstudeeropdracht richt zich op het verder ontwikkelen van Ricgraph door het onderzoeken en toepassen van AI/LLM-technieken voor automatische topicgeneratie, gecombineerd met de realisatie van een gebruikersinterface-koppeling om dit resultaat gericht te kunnen testen. Het project onderzoekt hoe de gegenereerde topics bruikbaar, reproduceerbaar en direct toetsbaar kunnen worden gemaakt, zodat onderzoeksinformatie efficiënter kan worden benut en nieuwe inzichten sneller beschikbaar komen.

## 2. Organisatorische context

De Universiteit Utrecht is een grote onderwijs- en onderzoeksinstelling met ruim 7.000 medewerkers. Binnen de universiteit is de directie Information and Technology Services (ITS) verantwoordelijk voor de ontwikkeling en het beheer van centrale ICT-diensten. ITS bestaat uit meerdere clusters, waaronder Applications.

Deze opdracht vindt plaats binnen de afdeling Research and Data Management Services (RDMS), zie Figuur 1 in de bijlage, onderdeel van het cluster Applications. RDMS houdt zich bezig met onderzoeksdata en onderzoeksinformatie. Een van de projecten binnen RDMS is Ricgraph (Research in context graph). Dit systeem brengt informatie over onderzoekers, publicaties, projecten en samenwerkingen samen in één graph. Daarbij worden ook nieuwe relaties afgeleid die in de afzonderlijke bronsystemen niet aanwezig zijn, waardoor verbanden en inzichten zichtbaar worden die anders verborgen zouden blijven.

De stage vindt plaats binnen dit project. De relevante ICT-aspecten voor deze opdracht zijn het datamodel van Ricgraph, de koppeling met de gebruikersinterface, en het gebruik van AI/LLM-modellen voor het verwerken en verrijken van onderzoeksinformatie.

## 3. Kwestie

Ricgraph is een flexibel systeem dat onderzoek informatie uit meerdere bronnen samenbrengt en verrijkt met nieuwe verbanden. Een van de uitbreidingen waar nu aan gewerkt wordt, richt zich op het automatisch genereren en toekennen van topics aan publicaties, datasets en software met behulp van AI en Large Language Models (LLM's). Dit zou onderzoekers, beleidsmakers en ondersteunend personeel in staat stellen sneller inzicht te krijgen in trends, expertise en samenwerkingen.

Op dit moment ontbreekt echter een betrouwbare methode om topics reproduceerbaar en bruikbaar te genereren én is er geen manier om de resultaten direct zichtbaar te maken in de Ricgraph gebruikersinterface. Hierdoor kunnen de gegenereerde topics nog niet praktisch worden getest of geëvalueerd voor eindgebruik. Dit belemmert de voortgang van de AI-uitbreiding en vergroot de kans dat deze onvoldoende aansluit op de behoeften van de praktijk. Wanneer dit niet wordt aangepakt, blijft veel van de potentiële waarde van Ricgraph onbenut.

## 4. Doel

Het doel van dit project is om de keten voor automatische topicgeneratie binnen Ricgraph te realiseren door een koppeling met de gebruikersinterface (UI) te ontwikkelen. Deze koppeling maakt het mogelijk om de resultaten van AI/LLM-modellen praktisch zichtbaar te maken en direct te testen binnen de context van Ricgraph.

Daarnaast wordt onderzocht welke AI-modellen en toepassingsmethodes het meest geschikt zijn om op een betrouwbare en consistente manier topics te genereren. Hierbij wordt geëxperimenteerd met verschillende modellen en configuraties. De best presterende aanpak wordt geselecteerd en geïntegreerd in de keten, zodat de resultaten direct bruikbaar zijn in Ricgraph.

Het project is geslaagd wanneer:

- Een user interface in Ricgraph Explorer is gerealiseerd die de gebruiker in staat stelt om een selectie van maximaal tien topics uit de bestaande Ricgraph-topics te maken, waarna (a) de bijbehorende research results worden weergegeven en (b) alle personen die aan deze research results gekoppeld zijn, zichtbaar worden. Hierbij geldt een logische AND-relatie tussen de gekozen topics.
- De geselecteerde topics afkomstig zijn uit een hiërarchische thesaurus van termen (broader & narrower). Voor deze opdracht worden de OpenAlex-concepten gebruikt (<https://docs.openalex.org/api-entities/concepts>), waarbij de door de LLM gegenereerde topics op hetzelfde hiërarchische niveau vallen. Er is gekozen voor level 2, met circa 21.000 concepten.
- De best presterende methode voor het genereren van topics is gerealiseerd en functioneel binnen de Ricgraph-keten.
- Een scriptie is opgeleverd waarin de onderzoeksresultaten en de onderbouwing van gemaakte keuzes uitgebreid zijn gedocumenteerd.
- Alle code en bijbehorende documentatie, inclusief de scriptie, op een goed gedocumenteerde manier beschikbaar zijn gesteld in een UU GitHub-repository.

## 5. Randvoorwaarden

Om een stage succesvol te laten verlopen, dienen aan een aantal randvoorwaarden te worden voldaan. Deze voorwaarden zijn van belang voor alle betrokken partijen: de student, de opleiding en de stagegever.

- Toegang tot gebouw en werkplek: De student heeft gedurende de normale werktijden toegang tot het gebouw en de werkplek van de organisatie.
- Toegang tot systemen: De student beschikt over de noodzakelijke systemen om de stagewerkzaamheden uit te voeren.
- Toegang tot netwerken: De student heeft toegang tot de relevante netwerken die benodigd zijn voor de uitvoering van de stageactiviteiten.
- Toegang tot werkaptop: De student kan gebruikmaken van een werkaptop indien dit nodig is voor de uitvoering van de stagewerkzaamheden.
- Toegang tot interviews: De student heeft de mogelijkheid om interviews af te nemen met diverse ontwikkelaars en andere medewerkers binnen de afdeling.
- Beschikbaarheid van de bedrijfsbegeleider: De bedrijfsbegeleider is beschikbaar voor regelmatig overleg met de student en kan op verzoek sporadisch overleg voeren met de stagebegeleider van de opleiding.

## 6. Afbakening

De scope van dit project wordt nadrukkelijk beperkt:

Processen: alleen de processen rondom het genereren en toekennen van topics worden onderzocht.

Andere AI-toepassingen (zoals chatbotfunctionaliteit) vallen buiten de scope.

Systemen: de focus ligt volledig op Ricgraph, het topics proces en de koppeling met de bestaande UI.

Andere onderzoeks-informatiesystemen worden enkel in context meegenomen.

Fasen: het project omvat onderzoek naar de beste methode voor de topics generatie en de realisatie van een werkende UI-koppeling om de resultaten zichtbaar te maken. Een volledige productieklare implementatie of bredere adviesopdracht maakt geen deel uit van dit project.

## 7. Producten

De producten van deze stage zijn een scriptie, een UI-koppeling en de implementatie van de best presterende methode.

Onderzoeksverslag (scriptie)

In de scriptie wordt het volledige onderzoeksproces vastgelegd. Dit omvat de analyse van verschillende AI/LLM-modellen en toepassingsmethodes, de uitvoering van experimenten, het vergelijken van resultaten en de onderbouwing van de gekozen aanpak.

UI-koppeling

De UI-koppeling verbindt de AI/LLM-topicgeneratie met de Ricgraph-gebruikersinterface. Hiermee worden de gegenereerde topics direct zichtbaar gemaakt, zodat verschillende methodes resultaat gericht kunnen worden geëvalueerd en vergeleken. Deze koppeling vormt de operationele basis voor het testen van de onderzoeksresultaten.

Implementatie van de best presterende methode

Op basis van de onderzoeksresultaten wordt de meest geschikte methode voor topicgeneratie geselecteerd en functioneel gerealiseerd binnen de Ricgraph-keten. Hiermee wordt een concreet, toepasbaar eindresultaat opgeleverd dat de bruikbaarheid van de onderzoeksuitkomsten aantoont.

## 8. ICT-component

Het onderzoek binnen deze opdracht richt zich op de toepassing van AI en valt daarmee binnen het ICT-vakgebied. De bevindingen zullen tevens praktisch worden getest. Daarnaast omvat het project een UI-koppeling als concreet beroepsproduct, waarbij technische vaardigheden, zoals het gebruik van Python, worden toegepast.

## 9. Beroepstaken

**Beroepstaak 1**

Taak	Analyseren
------	------------

<b>Architectuurlaag</b>	Organisatieprocessen
<b>Complexiteit context</b> 1□□☒3	Dit blijkt uit het feit dat Ricgraph een onderzoeksbreed systeem is dat data uit verschillende bronsystemen combineert. Stakeholders zijn zowel onderzoekers, beleidsmakers als externe partijen. Het project wordt uitgevoerd binnen een klein en dynamisch onderzoeksteam, waardoor doelen en prioriteiten sneller kunnen wisselen en voorspelbaarheid lager is. Daarnaast ontwikkelt AI/LLM-technologie zich snel, wat de complexiteit verder vergroot.
<b>Complexiteit taak</b> 1□□☒3	Dit blijkt uit de moeilijkheidsgraad van de opdracht: het onderzoeken, vergelijken en testen van AI/LLM-modellen en het implementeren van een UI-koppeling. Hiervoor is kennis nodig van data-analyse, AI, informatievoorziening, en systeemintegratie. Daarnaast vraagt de taak om het formuleren van onderbouwde keuzes en het uitvoeren van experimenten met reële onderzoeksdata.
<b>Zelfstandigheid</b> 1□☒□3	Dit blijkt uit het feit dat er grotendeels zelfstandig gewerkt zal worden, maar waar nodig wel hulp en inhoudelijke begeleiding kan krijgen van bedrijfsbegeleider. De verantwoordelijk ligt bij de student voor eigen voortgang en aanpak, maar kan terugvallen op begeleiding bij complexere vraagstukken.
<b>Totale beheersingsniveau beroepstaak</b>	1□□☒3
<b>Beroepstaak 2</b>	
<b>Taak</b>	Ontwerpen
<b>Architectuurlaag</b>	Organisatieprocessen
<b>Complexiteit context</b> 1□☒□3	Dit blijkt uit het feit dat het ontwerp plaatsvindt binnen Ricgraph, een relatief klein onderzoeksproject. Het team is dynamisch, wat soms leidt tot wisselende prioriteiten, maar de scope blijft beperkt tot de koppeling van AI/LLM-output naar de bestaande UI.
<b>Complexiteit taak</b> 1□□☒3	Dit blijkt uit het ontwerpen van een werkende koppeling waarmee de AI/LLM-output zichtbaar wordt in de Ricgraph-gebruikersinterface. Dit vereist kennis van datamodels, basisarchitectuur en integratie met bestaande systemen. Het is geen volledig strategisch ontwerp of architectuurwijziging, maar wel meer dan een eenvoudig implementatieplan.
<b>Zelfstandigheid</b> 1□☒□3	Dit blijkt uit het feit dat de opdracht grotendeels zelfstandig uitgevoerd zal worden, maar waar nodig wel terug kan vallen op bedrijfsbegeleider voor inhoudelijke en technische begeleiding.
<b>Totale beheersingsniveau beroepstaak</b>	1□☒□3

## 10. Aantonen onderzoekend vermogen

Deze afstudeeropdracht is geschikt omdat de nadruk ligt op het systematisch onderzoeken en evalueren van AI/LLM-technieken voor topicgeneratie binnen Ricgraph. Het gaat niet alleen om bouwen, maar vooral om onderbouwd experimenteren en analyseren.

De kenniscomponent zit in het bestuderen en toepassen van literatuur en bestaande onderzoeken over reproduceerbaarheid van AI-resultaten, topicgeneratie en evaluatiemethoden. Hiermee worden meetcriteria en tests om verschillende modellen en configuraties te vergelijken ontwikkeld. Scripties of artikelen over topic modeling, LLM-evaluaties en informatiebeheer met AI zijn hierbij relevante bronnen.

Het beroepsproduct bestaat uit een scriptie waarin de onderzoeksresultaten en gemaakte keuzes worden vastgelegd. De UI-koppeling dient daarbij als praktisch middel om de output zichtbaar en toetsbaar te maken.

De resultaten zijn ook bruikbaar voor anderen, bijvoorbeeld onderzoekers of studenten die AI-toepassingen willen testen. Zij kunnen profiteren van de beschreven methodes om topics reproduceerbaar en betrouwbaar te genereren.

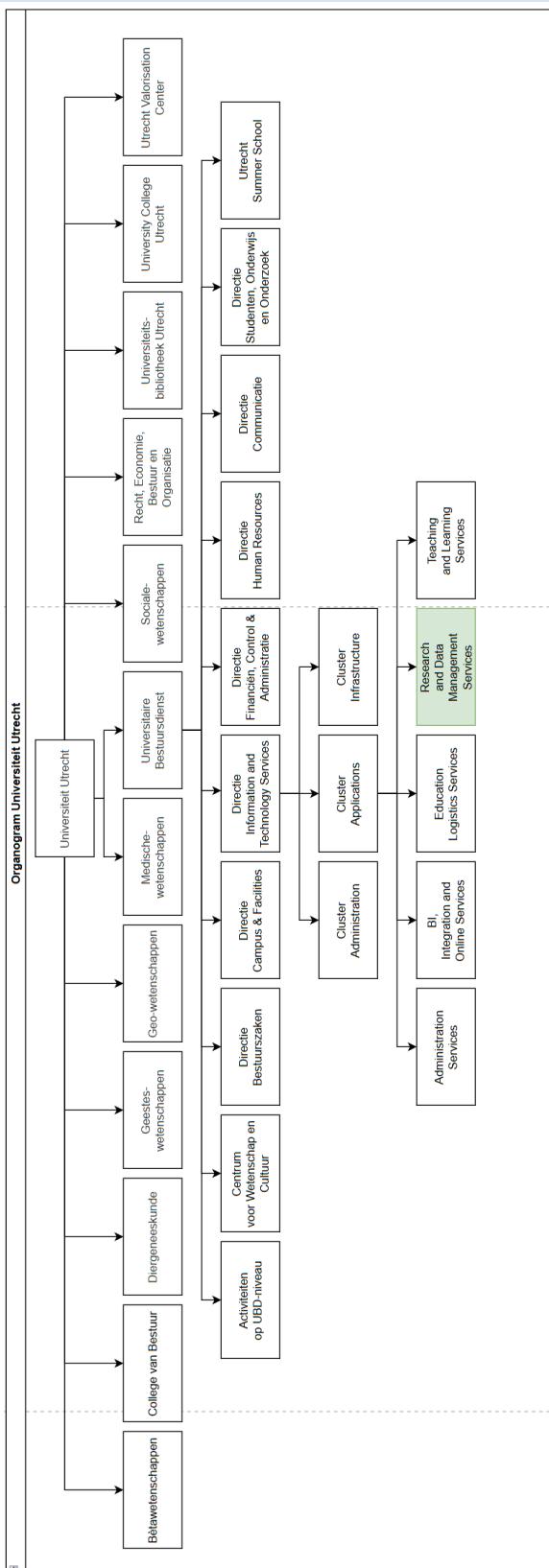
## 11. Ethische afwegingen

Het gebruik van AI/LLM's brengt meerdere ethische aandachtspunten met zich mee, zoals hallucinaties, beperkte transparantie en mogelijke bias in de resultaten. Ook is er een klimaatimpact door hoge energieconsumptie. Om deze risico's te beperken wordt uitsluitend gewerkt met toegestane, niet-persoonsgebonden data. De betrouwbaarheid, consistentie en mogelijke bias van de resultaten worden getest en keuzes en beperkingen transparant gedocumenteerd. Modellen worden efficiënt ingezet om de impact op het klimaat te beperken. Bij alle ethische kwesties vindt overleg plaats met de bedrijfsbegeleider.

## 12. Niveau bedrijfsbegeleider

De bedrijfsbegeleider, [...], is [...] en werkt als [...] op de universiteit. Dit betekent dat deze geschikt is om op hbo niveau te begeleiden. Daarnaast zal dr. Rik Janssen (Informatiemanager) dicht betrokken zijn bij het project en beschikbaar zijn om bijvoorbeeld rollen als opdrachtgever op te pakken.

## 13. Bijlage



Figuur 1 - Organogram Universiteit Utrecht