

# Titel van de bachelorproef.

## Optionele ondertitel.

---

**Ernst Aarden.**

Scriptie voorgedragen tot het bekomen van de graad van  
Professionele bachelor in de toegepaste informatica

**Promotor:** Dhr. F. Van Houte

**Co-promotor:** Mevr. S. Beeckman

**Academiejaar:** 2024–2025

**Eerste examenperiode**

**Departement IT en Digitale Innovatie .**

**HO  
GENT**



# Woord vooraf

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

# Samenvatting

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

# Inhoudsopgave

<b>Lijst van figuren</b>	<b>vi</b>
<b>Lijst van tabellen</b>	<b>vii</b>
<b>Lijst van codefragmenten</b>	<b>viii</b>
<b>1 Inleiding</b>	<b>1</b>
1.1 Probleemstelling . . . . .	1
1.2 Onderzoeksvraag . . . . .	2
1.3 Onderzoeksdoelstelling . . . . .	2
1.4 Opzet van deze bachelorproef . . . . .	2
<b>2 Stand van zaken</b>	<b>3</b>
<b>3 Methodologie</b>	<b>8</b>
<b>4 Conclusie</b>	<b>10</b>
<b>A Onderzoeksvoorstel</b>	<b>12</b>
A.1 Inleiding . . . . .	12
A.2 Literatuurstudie . . . . .	14
A.3 Methodologie . . . . .	15
A.4 Verwacht resultaat, conclusie . . . . .	17

# Lijst van figuren

2.1 Voorbeeld figuur. . . . .	4
A.1 Chronologische oplistng van de uit te voeren stappen . . . . .	13

# Lijst van tabellen

2.1 Voorbeeld tabel . . . . .	7
-------------------------------	---

# Lijst van codefragmenten

2.1	Voorbeeld codefragment . . . . .	4
-----	----------------------------------	---



# 1

## Inleiding

De inleiding moet de lezer net genoeg informatie verschaffen om het onderwerp te begrijpen en in te zien waarom de onderzoeksvraag de moeite waard is om te onderzoeken. In de inleiding ga je literatuurverwijzingen beperken, zodat de tekst vlot leesbaar blijft. Je kan de inleiding verder onderverdelen in secties als dit de tekst verduidelijkt. Zaken die aan bod kunnen komen in de inleiding (**Pollefliet2011**):

- context, achtergrond
- afbakenen van het onderwerp
- verantwoording van het onderwerp, methodologie
- probleemstelling
- onderzoeksdoelstelling
- onderzoeksvraag
- ...

### 1.1. Probleemstelling

Uit je probleemstelling moet duidelijk zijn dat je onderzoek een meerwaarde heeft voor een concrete doelgroep. De doelgroep moet goed gedefinieerd en afgeleid zijn. Doelgroepen als “bedrijven,” “KMO’s”, systeembeheerders, enz. zijn nog te vaag. Als je een lijstje kan maken van de personen/organisaties die een meerwaarde zullen vinden in deze bachelorproef (dit is eigenlijk je steekproefkader), dan is dat een indicatie dat de doelgroep goed gedefinieerd is. Dit kan een enkel bedrijf zijn of zelfs één persoon (je co-promotor/opdrachtgever).

## 1.2. Onderzoeksvraag

Wees zo concreet mogelijk bij het formuleren van je onderzoeksvraag. Een onderzoeksvraag is trouwens iets waar nog niemand op dit moment een antwoord heeft (voor zover je kan nagaan). Het opzoeken van bestaande informatie (bv. “welke tools bestaan er voor deze toepassing?”) is dus geen onderzoeksvraag. Je kan de onderzoeksvraag verder specificeren in deelvragen. Bv. als je onderzoek gaat over performantiemetingen, dan

## 1.3. Onderzoeksdoelstelling

Wat is het beoogde resultaat van je bachelorproef? Wat zijn de criteria voor succes? Beschrijf die zo concreet mogelijk. Gaat het bv. om een proof-of-concept, een prototype, een verslag met aanbevelingen, een vergelijkende studie, enz.

## 1.4. Opzet van deze bachelorproef

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt een overzicht gegeven van de stand van zaken binnen het onderzoeksdomein, op basis van een literatuurstudie.

In Hoofdstuk 3 wordt de methodologie toegelicht en worden de gebruikte onderzoekstechnieken besproken om een antwoord te kunnen formuleren op de onderzoeksvragen.

In Hoofdstuk 4, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeksvragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

# 2

## Stand van zaken

Dit hoofdstuk bevat je literatuurstudie. De inhoud gaat verder op de inleiding, maar zal het onderwerp van de bachelorproef \*diepgaand\* uitspitten. De bedoeling is dat de lezer na lezing van dit hoofdstuk helemaal op de hoogte is van de huidige stand van zaken (state-of-the-art) in het onderzoeksdomein. Iemand die niet vertrouwd is met het onderwerp, weet nu voldoende om de rest van het verhaal te kunnen volgen, zonder dat die er nog andere informatie moet over opzoeken (**Pollefliet2011**).

Je verwijst bij elke bewering die je doet, vakterm die je introduceert, enz. naar je bronnen. In  $\text{\LaTeX}$  kan dat met het commando `\textcite{}` of `\autocite{}`. Als argument van het commando geef je de “sleutel” van een “record” in een bibliografische databank in het Bib $\text{\LaTeX}$ -formaat (een tekstbestand). Als je expliciet naar de auteur verwijst in de zin (narratieve referentie), gebruik je `\textcite{}`. Soms is de auteursnaam niet expliciet een onderdeel van de zin, dan gebruik je `\autocite{}` (referentie tussen haakjes). Dit gebruik je bv. bij een citaat, of om in het bijschrift van een overgenomen afbeelding, broncode, tabel, enz. te verwijzen naar de bron. In de volgende paragraaf een voorbeeld van elk.

**Knuth1998****<empty citation>** schreef een van de standaardwerken over sorteer- en zoekalgoritmen. Experts zijn het erover eens dat cloud computing een interessante opportuniteit vormen, zowel voor gebruikers als voor dienstverleners op vlak van informatietechnologie (**Creager2009**).

Let er ook op: het `cite`-commando voor de punt, dus binnen de zin. Je verwijst meteen naar een bron in de eerste zin die erop gebaseerd is, dus niet pas op het einde van een paragraaf.

Sed commodum posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis



**Figuur 2.1:** Voorbeeld van invoegen van een figuur. Zorg altijd voor een uitgebreid bijschrift dat de figuur volledig beschrijft zonder in de tekst te moeten gaan zoeken. Vergeet ook je bronvermelding niet!

---

```
1 import pandas as pd
2 import seaborn as sns
3
4 penguins = sns.load_dataset('penguins')
5 sns.relplot(data=penguins, x="flipper_length_mm",
              y="bill_length_mm", hue="species")
```

---

**Codefragment 2.1:** Voorbeeld van het invoegen van een codefragment.

dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed

dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl. Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

Nulla mattis luctus nulla. Duis commodo velit at leo. Aliquam vulputate magna et leo. Nam vestibulum ullamcorper leo. Vestibulum condimentum rutrum mauris. Donec id mauris. Morbi molestie justo et pede. Vivamus eget turpis sed nisl cursus tempor. Curabitur mollis sapien condimentum nunc. In wisi nisl, malesuada at, dignissim sit amet, lobortis in, odio. Aenean consequat arcu a ante. Pellentesque porta elit sit amet orci. Etiam at turpis nec elit ultricies imperdiet. Nulla facilisi. In hac habitasse platea dictumst. Suspendisse viverra aliquam risus. Nullam pede justo, molestie nonummy, scelerisque eu, facilisis vel, arcu.

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

Kolom 1	Kolom 2	Kolom 3
$\alpha$	$\beta$	$\gamma$
A	10.230	a
B	45.678	b
C	99.987	c

**Tabel 2.1:** Voorbeeld van een tabel.

Donec et nisl at wisi luctus bibendum. Nam interdum tellus ac libero. Sed sem justo, laoreet vitae, fringilla at, adipiscing ut, nibh. Maecenas non sem quis tortor eleifend fermentum. Etiam id tortor ac mauris porta vulputate. Integer porta neque vitae massa. Maecenas tempus libero a libero posuere dictum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean quis mauris sed elit commodo placerat. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Vivamus rhoncus tincidunt libero. Etiam elementum pretium justo. Vivamus est. Morbi a tellus eget pede tristique commodo. Nulla nisl. Vestibulum sed nisl eu sapien cursus rutrum.

Nulla non mauris vitae wisi posuere convallis. Sed eu nulla nec eros scelerisque pharetra. Nullam varius. Etiam dignissim elementum metus. Vestibulum faucibus, metus sit amet mattis rhoncus, sapien dui laoreet odio, nec ultricies nibh augue a enim. Fusce in ligula. Quisque at magna et nulla commodo consequat. Proin accumsan imperdiet sem. Nunc porta. Donec feugiat mi at justo. Phasellus facilisis ipsum quis ante. In ac elit eget ipsum pharetra faucibus. Maecenas viverra nulla in massa.

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

# 3

## Methodologie

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt congue.

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

Maecenas non massa. Vestibulum pharetra nulla at lorem. Duis quis quam id lacus dapibus interdum. Nulla lorem. Donec ut ante quis dolor bibendum condimentum. Etiam egestas tortor vitae lacus. Praesent cursus. Mauris bibendum pede at elit. Morbi et felis a lectus interdum facilisis. Sed suscipit gravida turpis. Nulla at



lectus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Praesent nonummy luctus nibh. Proin turpis nunc, congue eu, egestas ut, fringilla at, tellus. In hac habitasse platea dictumst.

Vivamus eu tellus sed tellus consequat suscipit. Nam orci orci, malesuada id, gravida nec, ultricies vitae, erat. Donec risus turpis, luctus sit amet, interdum quis, porta sed, ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempor orci, et tincidunt urna nunc a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed vitae sem. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Maecenas ante. Duis ullamcorper enim. Donec tristique enim eu leo. Nullam molestie elit eu dolor. Nullam bibendum, turpis vitae tristique gravida, quam sapien tempor lectus, quis pretium tellus purus ac quam. Nulla facilisi.

# 4

## Conclusie

Curabitur nunc magna, posuere eget, venenatis eu, vehicula ac, velit. Aenean ornare, massa a accumsan pulvinar, quam lorem laoreet purus, eu sodales magna risus molestie lorem. Nunc erat velit, hendrerit quis, malesuada ut, aliquam vitae, wisi. Sed posuere. Suspendisse ipsum arcu, scelerisque nec, aliquam eu, molestie tincidunt, justo. Phasellus iaculis. Sed posuere lorem non ipsum. Pellentesque dapibus. Suspendisse quam libero, laoreet a, tincidunt eget, consequat at, est. Nullam ut lectus non enim consequat facilisis. Mauris leo. Quisque pede ligula, auctor vel, pellentesque vel, posuere id, turpis. Cras ipsum sem, cursus et, facilisis ut, tempus euismod, quam. Suspendisse tristique dolor eu orci. Mauris mattis. Aenean semper. Vivamus tortor magna, facilisis id, varius mattis, hendrerit in, justo. Integer purus.

Vivamus adipiscing. Curabitur imperdiet tempus turpis. Vivamus sapien dolor, congue venenatis, euismod eget, porta rhoncus, magna. Proin condimentum pretium enim. Fusce fringilla, libero et venenatis facilisis, eros enim cursus arcu, vitae facilisis odio augue vitae orci. Aliquam varius nibh ut odio. Sed condimentum condimentum nunc. Pellentesque eget massa. Pellentesque quis mauris. Donec ut ligula ac pede pulvinar lobortis. Pellentesque euismod. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent elit. Ut laoreet ornare est. Phasellus gravida vulputate nulla. Donec sit amet arcu ut sem tempor malesuada. Praesent hendrerit augue in urna. Proin enim ante, ornare vel, consequat ut, blandit in, justo. Donec felis elit, dignissim sed, sagittis ut, ullamcorper a, nulla. Aenean pharetra vulputate odio.

Quisque enim. Proin velit neque, tristique eu, eleifend eget, vestibulum nec, lacus. Vivamus odio. Duis odio urna, vehicula in, elementum aliquam, aliquet laoreet, tellus. Sed velit. Sed vel mi ac elit aliquet interdum. Etiam sapien neque, convallis et, aliquet vel, auctor non, arcu. Aliquam suscipit aliquam lectus. Proin tincidunt magna sed wisi. Integer blandit lacus ut lorem. Sed luctus justo sed enim.

Morbi malesuada hendrerit dui. Nunc mauris leo, dapibus sit amet, vestibulum et, commodo id, est. Pellentesque purus. Pellentesque tristique, nunc ac pulvinar adipiscing, justo eros consequat lectus, sit amet posuere lectus neque vel augue. Cras consectetur libero ac eros. Ut eget massa. Fusce sit amet enim eleifend sem dictum auctor. In eget risus luctus wisi convallis pulvinar. Vivamus sapien risus, tempor in, viverra in, aliquet pellentesque, eros. Aliquam euismod libero a sem. Nunc velit augue, scelerisque dignissim, lobortis et, aliquam in, risus. In eu eros. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Curabitur vulputate elit viverra augue. Mauris fringilla, tortor sit amet malesuada mollis, sapien mi dapibus odio, ac imperdiet ligula enim eget nisl. Quisque vitae pede a pede aliquet suscipit. Phasellus tellus pede, viverra vestibulum, gravida id, laoreet in, justo. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer commodo luctus lectus. Mauris justo. Duis varius eros. Sed quam. Cras lacus eros, rutrum eget, varius quis, convallis iaculis, velit. Mauris imperdiet, metus at tristique venenatis, purus neque pellentesque mauris, a ultrices elit lacus nec tortor. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent malesuada. Nam lacus lectus, auctor sit amet, malesuada vel, elementum eget, metus. Duis neque pede, facilisis eget, egestas elementum, nonummy id, neque.



# Onderzoeksvoorstel

Het onderwerp van deze bachelorproef is gebaseerd op een onderzoeksvoorstel dat vooraf werd beoordeeld door de promotor. Dat voorstel is opgenomen in deze bijlage.

## A.1. Inleiding

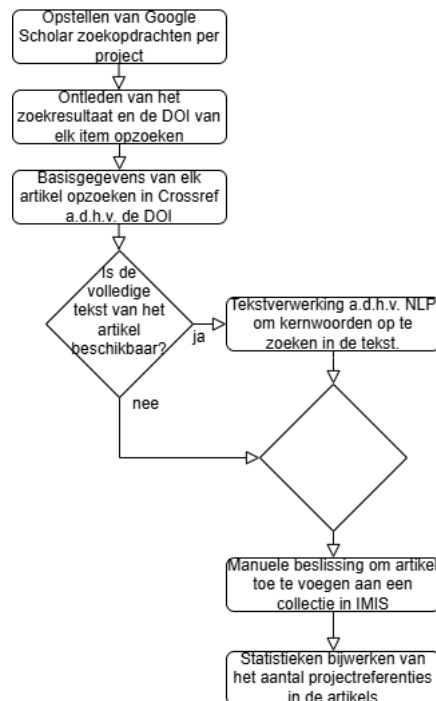
Het Vlaams Instituut voor de Zee (VLIZ) (**Vliz2024**) is een pionier in zeekennis. Dit wetenschappelijk instituut gelegen in Oostende heeft onder andere een mandaat om een complete en geactualiseerde catalogus bij te houden van alle wetenschappelijke publicaties in de mariene sector. Al meer dan 20 jaar bouwt het Integrated Marine Information System (IMIS) aan deze catalogus die intussen beschikt over meerdere collecties van marien wetenschappelijk referentiemateriaal.

Naast wetenschappelijke literatuur zitten er ook collecties van mariene wetenschappelijke projecten in het systeem. Het is belangrijk voor het VLIZ om te weten door hoeveel publicaties er naar een project (vb. het World Register of Marine Species (WoRMS) (**Worms2024**)) verwezen wordt. Deze maatstaf geeft een indicatie van het draagvlak van elk project binnen de wetenschappelijke gemeenschap en is één van de belangrijkste criteria tijdens projectevaluaties. Daarom is het cruciaal om continu nieuwe publicaties waarin verwezen wordt naar die projecten op te zoeken. Daarvoor wordt op heden Google Scholar gebruikt die bekend staat als de meest uitgebreide en geactualiseerde index. Op die manier blijft IMIS up-to-date en zijn de projectreferenties steeds geactualiseerd.

Momenteel verloopt dit proces binnen het VLIZ grotendeels handmatig. De zoekresultaten, volgens een bepaalde zoekfilter per project, afkomstig van Google Scholar worden manueel gefilterd en de Digital Object Identifier (DOI) van de geselecteerde artikels wordt opgezocht. Vervolgens worden de basisgegevens van elk artikel zoals titel, auteurs, datum en uitgever opgevraagd op basis van de DOI in Cross-

ref (**Crossref2024**). Met deze informatie wordt manueel beslist om het artikel al dan niet toe te voegen aan een collectie binnen IMIS.

Dit is een tijdrovend proces. Daarom is er vraag naar automatisatie die de zoekresultaten verwerkt en gestructureerd opslaat. De beslissing om een artikel toe te voegen aan IMIS blijft nog altijd een manuele stap, maar naar verwachting moet dit sneller, efficiënter en accurater verlopen. Dat moet mogelijk zijn doordat de heterogene zoekresultaten omgezet worden in gestructureerde informatie.



**Figuur A.1:** Chronologische oplijsting van de uit te voeren stappen

De centrale vraag die opgelost moet worden is: 'Hoe kunnen de uiteenlopende zoekresultaten van Google Scholar automatisch verwerkt worden om de DOI van elk item op te zoeken?'. Daarbij moet ook rekening gehouden worden met duplicaten, relevantie en diverse media types voor elk item van het zoekresultaat.

Om bovenstaande vraag te verduidelijken, moeten er een paar deelvragen gesteld worden:

- De zoekresultaten zijn afhankelijk van de zoekopdracht. In welke mate beïnvloeden kleine variaties van de zoekopdracht de resultaten?
- De zoekresultaten worden standaard gesorteerd op relevantie door Google Scholar, maar komt dat overeen met onze verwachtingen?
- De zoekopdracht kan het aantal resultaten beperken naar wens. Wat is de optimale threshold?

Om bovenstaande vraag op te lossen, moeten er een paar deelvragen beantwoord worden:

- Op welke manier kan de Google Scholar zoekopdracht op regelmatige basis automatisch uitgevoerd worden?
- Hoe worden de zoekresultaten van Google Scholar het best geparsed?
- Zoals gesteld zijn de zoekresultaten divers van aard. Kan het resultaat mits kwaliteitscontrole verbeterd (gefilterd) worden?
- De DOI zelf ontbreekt in het resultaat. Hoe zal de DOI voor elk item opgezocht worden?

Verder moet er gekeken worden welke databanktechnologie het meest geschikt is om de verwerkte informatie van elk item op te slaan. Tenslotte moet er ook nagedacht worden hoe het manuele proces met gebruik van de gestructureerde informatie het best softwarematig ondersteund kan worden. De volledige pijplijn zal geïmplementeerd worden als een Proof of Concept (PoC) die kan getest worden op basis van echte zoekresultaten. Als vervolg op dit onderzoek kan dan overgegaan worden tot het aanpassen en uitrollen van het systeem in functie van de feedback.

## A.2. Literatuurstudie

IMIS (**Haspeslagh2024**) heeft als doel het mariene onderzoek in Vlaanderen te coördineren. Die rol wordt vervuld door op te treden als centraal kenniscentrum voor en door de mariene sector in Vlaanderen. De taken van IMIS gaan verder dan louter het uitbouwen van collecties met publicaties, ze omvatten ook referenties naar wetenschappers en naar projecten. In deze bachelorproef zijn de publicaties en de wetenschappelijke projecten belangrijk.

Google Scholar is een zoekmachine die wetenschappelijke literatuur indexeert (**Wikipedia2024**). Google Scholar laat toe om te gaan zoeken op sleutelwoorden, titel, auteur, domein en combinaties van deze (**Noruzi2005**). Op die manier zijn er enkel zoekresultaten die overeenkomen met onze zoekcriteria.

Er zijn verschillende kenmerken die bijdragen aan het succes van Google Scholar ten opzichte van de andere grote collecties. Google Scholar is beschikbaar zonder kosten, het bevat de grootste bibliografische collectie ter wereld, en de data is afkomstig van zowel publieke als niet publieke bronnen (**Aguillo2011**).

Google Scholar wordt meer en meer gebruikt als een bibliometrische tool om informatie te verzamelen over de impact van citaties van en naar afzonderlijke artikels (**Moed2016**). **Araujo2021<empty citation>** legt uit hoe wetenschappelijke artikels andere publicaties citeren. Dat doen ze door middel van bibliografische referenties naar andere documenten doorheen de tekst, en een uitgewerkte lijst van referenties aan het einde van het artikel. Een index van citaties zoals in Google Scholar, is een databank die deze referenties tussen documenten opslaat.

In dergelijk digitaal netwerk van referenties, moeten de afzonderlijke entiteiten blijvend, betrouwbaar en onderscheidbaar geïdentificeerd kunnen worden. **Chandrakar2006<empty**

legt uit hoe de DOI voldoet aan deze criteria en de standaard identificatie is geworden van intellectuele eigendom op het internet.

Aan de hand van de bekomen DOI op basis van de zoekresultaten kan alle informatie met betrekking tot het artikel opgevraagd worden. In sommige gevallen is zelfs een pdf-versie beschikbaar. Vanuit ons standpunt zijn pdf-versies van wetenschappelijke artikels een vorm van ongestructureerde informatie. Automatisch berekende voorstellingen van de inhoud kunnen bijdragen om het zoekproces te vereenvoudigen. Dat is ook het uitgangspunt van het werk van **(Jehangir2023)**. Natural language processing (NLP) is van groot belang voor het verwerken van deze ongestructureerde data. Met behulp van NLP kunnen er conclusies en samenvattingen gemaakt worden. Binnen het domein van NLP, staat Named Entity Recognition (NER) centraal voor het opstellen van structuur op basis van ongestructureerde tekst **(Pakhale2023)**. **Palshikar**<empty citation> en **Jehangir2023**<empty citation> lijsten de verschillende NER technieken op die kunnen gebruikt worden. **Koning2005**<empty citation> tenslotte past NER toe om taxonomische entiteiten op te zoeken in een doorlopende tekst. Dit lijkt dusdanig ook geschikt voor het opzoeken van projectentiteiten in een artikel.

**Schaefer2011**<empty citation> merkt verder ook terecht op dat wetenschappers vandaag geconfronteerd worden met een dagelijkse vloedgolf van nieuwe publicaties. Daarenboven worden ook meer en meer oude artikels gedigitaliseerd. Zowel de verwerking van de platte tekst als de vloedgolf aan publicaties geven aanleiding tot het toepassen van NLP voor het beheren van wetenschappelijke artikels **(Schaefer2011)**.

### A.3. Methodologie

Ons onderzoek wordt opgesplitst in 7 afzonderlijke fases:

- opstellen van de meest accurate zoekopdrachten in Google Scholar
- parsen van de zoekresultaten om de DOI van elk artikel op te zoeken
- basisgegevens van elke publicatie opvragen in Crossref op basis van de DOI
- ontwikkelen van een tekstverwerkingsproces op basis van NER
- ontwerpen van de databank voor de opslag van de gestructureerde data
- installeren van een pijplijn die de afzonderlijke stappen combineert
- ontwerpen van een gebruikersinterface die de gestructureerde data toont

Achtereenvolgend wordt er nu dieper ingegaan op de verschillende fases. In de context van dit voorstel, wordt elke stap nog niet volledig uitgewerkt, maar is het de bedoeling om een richting aan te geven. Een uitvoerige beschrijving van alle taken zal aan bod komen in de bachelorproef.

De eerste fase gaat over het correct opstellen van de Google Scholar zoekopdrachten voor de verschillende projecten met als doel om als resultaat de meest relevante artikels te verkrijgen.

De tweede fase komt overeen met de hoofdpdracht van ons onderzoek, namelijk een systeem ontwikkelen dat de heterogene zoekresultaten kan parsen om de DOI op te zoeken. Meerdere zaken komen hier aan bod zoals ten eerste het aggregeren van zoekresultaten en groeperen per titel. Vervolgens het opzoeken van de DOI via de linked data, of via Crossref, of op de full-text versie van het artikel. Python lijkt de voor de hand liggende programmeertaal voor deze taak waarbij het draait om dataverwerking.

In de derde fase zal er met de API van Crossref gewerkt worden om de basisgegevens van een artikel op te vragen aan de hand van de DOI. Ook hier zal Python gebruikt worden.

De vierde fase onderzoekt of er met behulp van NER bruikbare informatie gefilterd kan worden uit de pdf-versie van een artikel. Op die manier kunnen de basisgegevens verrijkt worden om de manuele moderatie accurater te maken. Een zoektocht op het internet levert snel een achttal oplossingen op die ruwe tekst verwerken tot gestructureerde informatie:

- ScienceParse (**ScienceParse2024**)
- ScienceParse API  
(**ScienceParseAPI2024**)
- Grobid (**Grobid2024**)
- Sci-pdf parser (**SciPdfParser2024**)
- Paper AI (**Paperai2024**)
- Paper ETL (**PaperEtl2024**)
- Delph-in (**Delphin2024**)
- TaxonGrab (**TaxonGrab2024**)

Om onze specifieke business logica te implementeren, wordt gewerkt met één oplossing, of een combinatie van meerdere oplossingen, eventueel met gebruik van een geschikte NER methode. Ook hier is Python de gekozen programmeertaal voor de dataverwerking.

De vijfde fase evalueert welke databanktechnologie het best geschikt is om onze gestructureerde data in op te slaan. Daar zijn bijvoorbeeld relationele, grafische en document store databanken.

De zesde fase focust zich op de automatisatie van het gehele proces: van zoekresultaat tot opgeslagen gestructureerde data. Er zal een pijplijn opgesteld worden



die de voortgang van elke stap observeert en de status van het hele proces update. Dit wordt gebouwd met Symfony rekening houdend met de technology stack binnen het VLIZ.

De zevende en laatste fase tenslotte moet ervoor zorgen dat alle voorgaande stappen ook opbrengen voor de eindgebruiker. De opgeslagen gestructureerde data wordt ontsloten via een API (API Platform) en een custom front-end (NextJS) applicatie zorgt ervoor dat deze overzichtelijk voorgesteld wordt.

#### **A.4. Verwacht resultaat, conclusie**

Naar verwachting zal deze pijplijn de verwerking van de Google Scholar zoekresultaten versnellen. Op die manier draagt ons onderzoek bij aan de continue uitbreiding van de collecties van IMIS. Er wordt ook verwacht dat het aantal projectreferenties beter gemonitord zal worden. Zo helpt ons onderzoek bij het beter in kaart van brengen van het draagvlak voor de wetenschappelijke projecten van het VLIZ.