

## Sprint retrospective, Milestone #2

User story	Task	Assignee/s	Estimated effort	Actual effort	Done	Notes
As an admin, I want to persist a sufficient amount of articles in the database, so that I will get an accurate similarity ratio when running my algorithm on it	#41	Vlad Nitu	5h	7h 7m	Completed	See MR <a href="#">!23</a> For a detailed description of the request. Challenges: Python absolute and relative imports when running the scripts, cleaning the database from Utrecht hogeschool and writing tests by mocking with `@patch`.
Research algorithm for computing similarity between documents and boost its speed	Relates to #12	Vlad Nitu	1h	3h	In-progress	Still researching in order to work along with @cciacu; improved running speed by querying the database in batches (instead of using for-loops) and multi-threading implementation, but it is still slow
Research Heroku web servers in order to deploy our app	-	Vlad Nitu + Rares Toader	2h (Vlad Nitu), 2h (Rares Toader)	2h(Vlad Nitu), 1h30m(Rares Toader)	In-progress	Vlad Nitu => After client meeting and listening to his proposal, I've researched <a href="#">Heroku</a> get-started tutorial to understand how we are going to migrate our codebase to their platform (so that we won't run our servers locally anymore) Rares Toader => I've looked into different ways in which we could

						<p>deploy our web application towards the end of the project, and the two options I glanced upon were <a href="#">Heroku</a> and <a href="#">Firebase</a>. Both offer ways in which we could deploy the app and also solution for cloud storage, so we will have to see and decide which to go for in the following sprints.</p>
Fully test frontend side, including main & secondary pages	!28	Rares Toader & Vlad Nitu	4h	1h (Vlad) + 4h (Rares) = 5h	Completed	<p>Vlad: Added `checkURL` text in order to achieve 100% branch coverage by following @dmicloiu 's framework for testing  Rares: Added tests for the Home page, and for other secondary pages, augmenting them so they could achieve full branch coverage. Other important components that were tested were CheckTwoTexts.js, Footer.js and services.js. Made certain refactors for some components, so they would become more reusable and easier to test. Also changed nyc to ignore some of the files when reporting coverage</p>

"/checkURL" unexpected behavior (overflow form & button) on mobile phone	#44	Vlad Nitu	1h30m	2h	In-review	Followed @cciacu's advice and stick to using `vh` and `vw` when possible (depending on the context, but here `px` was not a desirable unit of measure). ShoThe page should be mobile-friendly now
As a user, I want to have a special option in the menu, so that I will have the ability to insert two pieces of text of news articles in two different text boxes.	#9	Matei Mirica	8h	5h30m	Completed	This issue is only related to creating the frontend page which displays two boxes. Some functionality was already implemented by @rtoader so it took less than expected to finish the work.
As a user, I want to check whether the two pieces of text that I have inputted are overlapping.	#14	Matei Mirica	10h	13h	In-review	This issue relates to making the 2 boxes functional: extracting the text from the boxes; send request to backend; compute similarity; return result to frontend; display result. It took more than expected because there were some issues with mocking the axios module and it was difficult to test.

As a user, I want to have a special option in the menu, so that I will have the ability to insert two URLs of news articles in two different text boxes.	#19	Diana Micloiu	12h	12h30m	Completed	This issue is only related to the frontend part and consists of creating a new page for checking the similarity between two URLs and modifying the main page - services section to also include the new service. In the merge request corresponding to this issue I also solved some inconsistencies regarding titles in the footer or design choices for the frontend. (mainly small details that needed to be fixed)
As a user, I want to check whether the two URLs that I have inputted are similar.	#20	Diana Micloiu	10h	14h	Completed	This issue relates to properly completing the functionality of the "Compare two URLs similarity" by implementing the backend API corresponding to this issue and then modifying the frontend to change accordingly to different interactions of the user. Nonetheless, it also includes making the connection between the two. Note that testing this frontend functionality proved to be challenging and that is the main reason behind the 4h hours unaccounted effort.

Researching and reading papers related to finding overlap between documents.	Relates to #12	Rares Toader	9h	6h 15m	In progress.	<p>As we had some trouble with the time efficiency of our algorithm, I looked further into the existing literature to find potential improvements. The papers I read are the following:</p> <p><a href="https://www.researchgate.net/profile/Raphael-Finkel/publication/221573993_Signature_Extraction_for_Overlap_Detection_in_Documents/links/0c96052af3e31f1ece000000/Signature-Extraction-for-Overlap-Detection-in-Documents.pdf">https://www.researchgate.net/profile/Raphael-Finkel/publication/221573993_Signature_Extraction_for_Overlap_Detection_in_Documents/links/0c96052af3e31f1ece000000/Signature-Extraction-for-Overlap-Detection-in-Documents.pdf</a></p> <p>=&gt; Signature Extraction for Overlap Detection (Melbourne)</p> <p><a href="https://citeseerx.ist.psu.edu/document?repid=rep1&amp;type=pdf&amp;doi=dc30cb8757ea5abcd8a4bea9bcf880aea4ee4696">https://citeseerx.ist.psu.edu/document?repid=rep1&amp;type=pdf&amp;doi=dc30cb8757ea5abcd8a4bea9bcf880aea4ee4696</a></p> <p>=&gt; The SCAM plagiarism detector from Stanford</p> <p><a href="https://citeseerx.ist.psu.edu/document?repid=rep1&amp;type=pdf&amp;doi=b7f0018949701d508c47706d30f9bf796fb5005e">https://citeseerx.ist.psu.edu/document?repid=rep1&amp;type=pdf&amp;doi=b7f0018949701d508c47706d30f9bf796fb5005e</a></p> <p>Koala system for detecting overlap from Bell Laboratories</p>
--	----------------	--------------	----	--------	--------------	---

Made the services option in the navbar to be a dropdown menu, with the corresponding connections.	#43	Rares Toader	2h	1h20m	Completed	In the navbar, the user can now hover over the services and can see a drop down menu, thus being able to choose any of the features to go to.
As a user, I want to enter the link of a news article, so that I will be given an overlap warning if the article has high similarity with one document in the database.	#12	Cristian Ciacu + Rares Toader	8h (Rares Toader) 20h (Ciacu Cristian)	28h	In progress	This issue refers to connecting the frontend button from the check url functionality to the backend. In the backend a query should happen where the fingerprint of the input url is compared to a subset of the news articles that we have stored. The current version of this algorithm is one that is too slow for any user to use. Rares => worked on implementing the connection between clicking the button and making the request to the backend. Also looked into possible ways in which the algorithm could be improved (deleting shingle_pos and using a subsample of the

					<p>hash values). Still need to implement strategy for error handling and helping out with the algorithm</p> <p>Cristian Ciacu =&gt; Tried to find ways of running a query to find similarity between inserted article and the articles from the database in a relatively decent amount of time. Some of the methods i've tried: 1 process (way to slow, around 400s a query), multi-threading, not much of an improvement, multi-processing (~65s with 64 processes (not a viable method, since it is very computationally expensive, ~120s with 32 processes, so basically it scales linearly). I've also tried sub-sampling(~10% from the submitted article fingerprints) and then perform multi-processing (16 processes =&gt; 25s a query)</p> <p>I've also taken a look into the similarity function, I've implemented the cosine similarity, and it gives surprisingly high similarity levels, but I</p>
--	--	--	--	--	--

						still need to investigate that.
--	--	--	--	--	--	---------------------------------