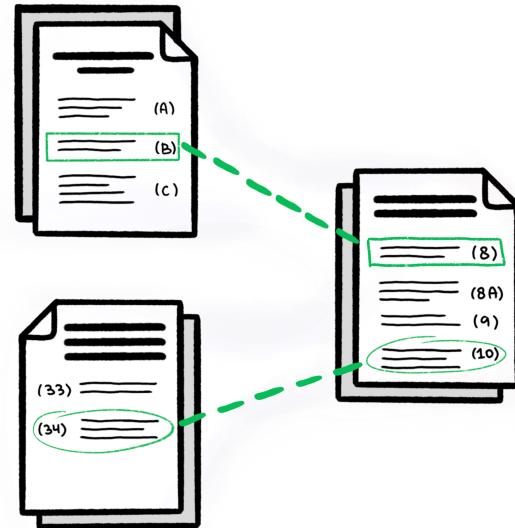




# A meta analysis of theoretical research

We live in what some have called the data age. In part this is reflected by the large amount of papers that are being published every year. Such papers can be found based on keywords or through citation networks. However, for more theoretically oriented work it is relevant not only whether a paper is topically related, but also whether it is methodologically related. A former master student, Mariia Vlasova, has pioneered work on connecting papers based on the equations that they use. She has collected all equations that were in papers in the journal of theoretical population biology from 2004 onwards. Based on this database she has identified papers that have identical formulas and made a promising start for finding fuzzy matches. The next step in this project is two-fold: increase the number of papers that are being compared and improve the fuzzy matching algorithm. In particular, a major challenge lies in representing the formulas in a way that 1) takes into account substitution through other equations in the same paper, 2) matches formulas even when they are written in a different way (e.g. compare 0.5 and  $\frac{1}{2}$ , although written differently, they typically have the same meaning). An important part of the project will concern exploring which existing tools might benefit our work. This project still has room to accommodate the individual interest of the student. Affinity and familiarity with coding in Python are a must for this project.



## Suggested reading

Farrell, M. J., Brierley, L., Willoughby, A., Yates, A., and Mideo, N. (2022). Past and future uses of text mining in ecology and evolution. *Proceedings of the Royal Society B*, 289(1975):20212721

Nunez-Mir, G. C., Iannone III, B. V., Pijanowski, B. C., Kong, N., and Fei, S. (2016). Automated content analysis: addressing the big literature challenge in ecology and evolution. *Methods in Ecology and Evolution*, 7(11):1262–1272

**Methods:** Python, data mining

**State of the project:** consolidating project

<b>Staff member:</b>	Koen van Benthem	<b>Contact:</b>	k.j.van.benthem@rug.nl
<b>Daily supervisor:</b>		<b>Contact:</b>	
<b>Expertise group:</b>	TRÈS		
<b>Type of project:</b>	<input type="checkbox"/> Bioinformatics	<input type="checkbox"/> Fieldwork	<input type="checkbox"/> Laboratory
<b>MSc program:</b>	<input checked="" type="checkbox"/> Biology	<input checked="" type="checkbox"/> Ecology and Evolution	<input checked="" type="checkbox"/> Theoretical
	<input type="checkbox"/> Biomedical Sciences	<input type="checkbox"/> Biomolecular Sciences	<input checked="" type="checkbox"/> Data analysis
<b>ECTS:</b>	<input checked="" type="checkbox"/> 30	<input checked="" type="checkbox"/> 40	<input checked="" type="checkbox"/> Behavioural and Cognitive Neurosciences
			<input type="checkbox"/> Marine Biology
		<b>Language:</b>	<input type="checkbox"/> Dutch
			<input checked="" type="checkbox"/> English
<b>Start date:</b>		<b>Location:</b>	



university of  
groningen

faculty of science  
and engineering



MSc Research Project