



**Data Science:  
Introduction to Text Mining with R (S41)  
19 – 22 August 2024**

Course Director: Dr. Ayoub Bagheri  
E-mail: [ms.summerschool@uu.nl](mailto:ms.summerschool@uu.nl)

Day	Time	Type	Description	Location
Monday	09:00 – 10:30	Lecture	Introduction to R & RStudio Introduction to text mining Regular expressions	<a href="#">Victor J. Koningsberger building</a> , room 224
	10:45 – 11:45	Computer Lab	R intro and processing textual data in R	
	11:45 – 12:30	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	
	12:30 – 14:00	Lunch Break	In <a href="#">Educatorium</a>	
	14:00 – 15:30	Lecture	Preprocessing text: <ul style="list-style-type: none"><li>- Cleaning</li><li>- Tokenization (n-grams)</li><li>- Stop words removal</li><li>- Stemming</li></ul>	
	15:45 – 16:30	Computer Lab	Text preprocessing Visualisation N-gram tokenization	
	16:30 – 17:00	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	



Day	Time	Type	Description	Location
<b>Tuesday</b>	09:00 – 10:30	Lecture	Text as features: - Vector space model - Representation Text classification	<a href="#">Victor J. Koningsberger building,</a> room 224
	10:45 – 11:45	Computer Lab	Training and test data Naïve Bayes and SVM	
	11:45 – 12:30	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	
	12:30 – 14:00	Lunch Break	In <a href="#">Educatorium</a>	
	14:00 – 15:30	Lecture	Sentiment analysis	
	15:45 – 16:30	Computer Lab	Sentiment classification	
	16:30 – 17:00	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	

Day	Time	Type	Description	Location
<b>Wednesday</b>	09:00 – 10:30	Lecture	Feature selection in text	<a href="#">Victor J. Koningsberger building,</a> room 224
	10:45 – 11:45	Computer Lab	Filter-based feature selection Embedded feature selection	
	11:45 – 12:30	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	
	12:30 – 14:00	Lunch Break	In <a href="#">Educatorium</a>	
	14:00 – 15:30	Lecture	Text clustering & topic modeling	
	15:45 – 16:30	Computer Lab	K-Means Latent Dirichlet allocation	
	16:30 – 17:00	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	



Day	Time	Type	Description	Location
Thursday	09:00 – 10:30	Lecture	Word embeddings	<a href="#">Victor J. Koningsberger building</a> , room 224
	10:45 – 11:45	Computer Lab	Training word vectors Vector similarities	
	11:45 – 12:30	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	
	12:30 – 14:00	Lunch Break	In <a href="#">Educatorium</a>	
	14:00 – 15:30	Lecture	Neural networks and text mining applications	
	15:45 – 16:30	Computer Lab	Using pre-trained word vectors Embedding layer	
	16:30 – 17:00	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	

For information about the Social Programme, please visit the [Utrecht Summer School website!](#)