

HEALTH TEXT PROCESSING WITH ORANGE

Erik Tjong Kim Sang¹, Peter Kok¹, Wouter Smink²,
Bernard Veldkamp³, Gerben Westerhof², Anneke Sools²

¹Netherlands eScience Center

²Psychology, Health and Technology, University of Twente

³Research Methodology, Measurement and Data Analysis, University of Twente

INTRODUCTION

Orange is a **free data mining toolkit** developed by the Bioinformatics Laboratory of the University of Ljubljana in Slovenia. The software offers a large collection of modules that can be combined in a **user-friendly graphical interface** to form arbitrary processing pipelines. Orange consists of open source Python code and can be extended with **custom-build modules**.

CASE STUDY

We used Orange for textual analysis of online correspondence between counselors and patients with substance dependence related health problems. We developed six project-specific Orange modules for data loading, data organization, data selection, data filtering, data analysis and data visualization.

RESULT

Orange allowed us to quickly assess various aspects of the online correspondence. The software is now used by both staff and students of the Psychology program of the University of Twente

in their research projects.

CONCLUDING REMARK

Orange is highly recommended for data analysis by people without experience in data mining or computer programming.

