

Data Mining: Automated Collection and Analysis of Textual & Numerical Data (AIDAHO)

Course access: https://ilias.uni-hohenheim.de/goto.php?target=crs_1572531&client_id=UHOH

Class Info

Time: Thursday 14:15-15:45
Location: [PC-Raum 2](#)
ECTS: 6

Instructor Info

Instructor: Prof. Dr. Jens Vogelgesang
E-Mail: j.vogelgesang@uni-hohenheim.de

Class description

Hop on the data digging adventure! This hands-on course is all about the nitty-gritty of using data science in real life, exploring the goldmine of data hidden on the internet and digital platforms. We're basically swimming in data that can answer old and brand-new questions. The tricky part? A lot of this data is kind of a hot mess or just barely put together. Take website content, like press releases, as an example. To analyze this, you need to get under the hood of the website's HTML code to find and extract the information you need. The same goes for PDFs, like annual reports, which might require a bit of optical character recognition magic (OCR) to turn images of text into actual data you can work with. Then there's the world of semi-structured data, served up through application programming interfaces (APIs). These APIs, like Spotify's Web API for fetching track features offer a more organized way to access data for analysis. Check out the example code below to get a feel for what we're going to do. It's like a sneak peek into grabbing data from somewhere cool like Spotify's API. If you are more interested in time series data, APIs also gain access to databases for stock market trends or biological process. Once we've scraped or pulled down the data, we're going to roll up our sleeves and get into statistical text analysis and number-crunching with stats to get to the bottom of our project questions.

Target Audience

This course is a part of the [AIDAHO study program](#) but it's wide open to Master's students from all disciplines. We'll be using R/RStudio and specific R libraries to unearth data from the web, PDFs, and APIs. If you've got experience with R/RStudio, great! If not, don't sweat it – your curiosity and willingness to learn are what count most.

Assessment and grades

You get to choose: fly solo or join forces with a buddy to dive into a research question you're curious about, using the cool data and methods we're going to explore together. Then, you'll put your discoveries on a scientific poster. Think of it as your research story on a big, colorful canvas. At the end of the semester, we're throwing a poster session where you'll get to show off your poster and talk everyone through what you found. It's a laid-back way to share your hard work and maybe even impress a few people!

You can apply to have this course credited as a free elective module (Freies Wahlmodul) or as a portfolio module.

Example Code

<https://www.rcharlie.com/spotify/>