

Baybars Külebi

gullabi.github.io

EDUCATION

Ruprecht-Karl-Universität Heidelberg

PH.D. IN ASTRONOMY

Sep 2007 – Dec 2010

Grade: Magna Cum Laude

Middle East Technical University

M.S. IN PHYSICS

Feb 2005 – Nov 2006

GPA: 3.50 / 4.00

B.S. IN PHYSICS

Oct 2002 – Jan 2005

GPA: 3.31 / 4.00

SKILLS

PROGRAMMING

Python:

Over 1M lines

C/C++:

Over 10k lines

Fortran 77/95:

Over 3k lines

DATA ARCHITECTURE

MongoDB • MySQL • PostgreSQL

DATA PROCESSING

SciPy • Matplotlib • scikit-learn

SPEECH RELATED

CMU-Sphinx • Kaldi • ARPA LM

WEB RELATED

flask • django • Celery

METHODOLOGIES

Agile • TDD • Microservices • API first

INFRASTRUCTURE

DigitalOcean • AWS • Docker • git • svn

MULTIMEDIA

Gimp • Audacity • kdenlive

LANGUAGES

Turkish Mother tongue

English, Catalan Fluent

Spanish Advanced

German Intermediate

LINKS

Bitbucket:// [bkulebi](#)

Github:// [gullabi](#)

Twitter:// [@gullabi](#)

NASA ADS:// [Baybars Kulebi](#)

DATA SCIENCE EXPERIENCE

TECHNOLOGY LEAD | DATA SCIENCE AND SPEECH ENGINEERING

April 2017 - Current | CollectivaT, Spain

- Text clustering for meaning using K-means and visualization of the results.
- Sequence alignment for translation assessment and bitext generation.
- Training the first open-source ASR system in Catalan, from 550 hours of acoustic data scraped from public resources.
- Training RNN based punctuation models in Keras.
- On premise pedestrian detection solution using OpenCV libraries and a raspberry pi.

PRODUCT ENGINEER | DEVELOPMENT IN MUSICAL RIGHTS DISTRIBUTION

March 2015 – July 2016 | BMAT, Spain

August 2016 – March 2017 | Freelance

- Was responsible for finding tailored solutions for the specific problems of each client and automatizing these processes.
- Data wrangling and enrichment for music rights management
- Audio fingerprinting and matching
- Audio similarity assessment and alignment using dynamic time warping

PH.D. | POST-DOC | SCIENTIFIC SOFTWARE DEVELOPMENT

Sep 2007 – Dec 2010 | Astronomisches Rechen-Institut, Germany

May 2011 – May 2014 | ICE-CSIC, Spain

- Data modeling using various **regression methods** for solving various research problems.
- Solving optimization problems with the use of **evolutionary algorithms**.

ADDITIONAL EXPERIENCE

TECHNOLOGY LEAD | MANAGEMENT OF OPERATIONS

April 2017 - Current | CollectivaT, Spain

- Development of a language-agnostic back-end system for pre-processing **Automated Speech Recognition (ASR)** training datasets.
- **System administration** of the digital infrastructure; servers and productivity tools.
- Automatizing twitter scraping process and setup a database to persist the data and facilitate the data analysis.