



# **Shrnutí celého vývoje projektu ze strany Západočeské univerzity v Plzni**

Pavel Ircing

Katedra kybernetiky  
Fakulta aplikovaných věd  
Západočeská univerzita v Plzni  
Plzeň

# Inspirace technická

Ačkoliv jsme ten „nejtechničtější tým“ v projektu, i my jsme se seznámili s formáty a technologiemi, které pro nás byly novinkou:

- formát WARC (Web ARChive)
- „framework“ Apache Hadoop a související nástroje
  - sada nástrojů pro práci s webovými archivy – Archives Unleashed Toolkit
- konkrétní metody pro efektivní práci s velkým objemem data – např. pro čištění HTML kódu



# Inspirace z oblasti společenských věd

- spolupráce se Sociologickým ústavem ukázala, že metody strojového učení jsou výzkumníky z oblasti společenských věd nadšeně vítány
  - např. automatická identifikace témat
- „společenskovědní“ týmy navíc v současnosti téměř vždy mají někoho, kdo se v matematicko-informatických záležitostech velmi dobře vyzná
  - + „demokratizace“ strojového učení
- (mimo jiné) na základě těchto zkušeností byl založen CCS-Lab



# Inspirace právní

- je třeba dávat pozor na ochranu autorských práv a osobních údajů
- na druhou stranu, je dobré, aby odborníci z oboru naopak „hlídali“ regulační orgány, aby nebylo strojové učení z dat regulací znemožněno
  - *Directive on Copyright in the Digital Single Market*
    - „copyright exceptions for text- and data-mining“
  - *EU Artificial Intelligence Act*
    - rámec pro regulaci všech typů umělé inteligence – momentálně připravovaný



Děkuji za pozornost!