Activity-based Search Session Identification
Samuel Molnár, Supervisor: Tomáš Kramár

Motivation
› Finding search goals
› Identifying similar queries
› Knowing user intent
› Recommending results
› Personalizing search

User Model
› User’s activity
› Shared results and their position
› Clicks on search results
› Dwell time on web pages

Method
› Time as feature
› Lexical features
› Our user model as feature
› Machine Learning model using SVM

Data & Results
› Our method is ~ 8% more accurate than widely used time methods
› Collected unique dataset with explicitly annotated sessions by users

30 users
464 sessions
2k queries
18k pages

Annotation Study
› Annotations perform worse than time methods
› Annotations depend on domain knowledge
› Mutual agreement on annotations is weak

Søke
› Our search engine
› Open Source project
› Platform for annotated search log acquisition
› Explicit annotations by users

github.com/smolnar/soke