

# Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** František Kmječ  
**Název práce** Methods of User-Assisted Summarization of Meetings  
**Rok odevzdání** 2023  
**Studijní program** Informatika  
**Specializace** Obecná informatika

**Autor posudku** doc. RNDr. Ondřej Bojar, Ph.D. Vedoucí  
**Pracoviště** Ústav formální a aplikované lingvistiky

## K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X			
Splnění zadání	X			
Rozsah práce <i>... textová i implementační část, zohlednění náročnosti</i>	X			

Bakalářská práce Františka Kmječe přesahuje běžný standard svým rozsahem i kvalitou zpracování a blíží se spíše práci diplomové. Práce představuje vlastní přístup k automatické sumarizaci zápisů z jednání a kompletní návrh, implementaci a praktické ověření nástroje pro kolektivní interaktivní vytváření zápisů během videohovoru. Návrh nového přístupu k sumarizaci dlouhých transkriptů (inkrementální přístup) empiricky testuje dva možné podkladové modely (BART a LED) a Llamu, velký jazykový model. Baseline se sice inkrementální metodou nedaří překonat, zejména z důvodu nedostatku trénovacích dat (dostupná data nemají dostatečně inkrementální zápisky), ale přístup zůstává nadějný a pečlivě popsaná pozorování jsou velmi užitečná. Vytvořený nástroj je prakticky použitelný a byl otestován i s neškolenými uživateli. Získaný přehled nejpálčivějších problémů sumarizace je velmi užitečný pro další výzkum.

Souběžně s bakalářskou prací František připravil *dva* vědecké články. Jednak popis svého systému do soutěže ve vytváření zápisů AutoMin 2023 (bude prezentováno v září na konferenci INLG v Praze), a jednak popis interaktivního nástroje jako demo paper (v recenzním řízení).

## Textová část práce

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Formální úprava <i>... jazyková úroveň, typografická úroveň, citace</i>	X			
Struktura textu <i>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>	X			
Analýza	X			
Vývojová dokumentace	X			
Uživatelská dokumentace	X	X		

Práce je psána velmi dobrou angličtinou, s jen velmi malým počtem chyb a překlepů (např. *unsuccessful*). V přehledné struktuře jsou popsány potřebné teoretické základy, modely a metody pro přepis mluvené řeči i pro sumarizaci. Analýza je velmi podrobně propracována z hlediska úlohy sumarizace, menší důraz je kladen na analýzu z hlediska SW inženýrství, ale to je pochopitelné, protože vytvořený nástroj má sloužit jako proof of concept, první prototyp. Dokumentace výsledného SW díla pro vývojáře je dostatečně podrobně poskytnuta v příloze 2. Uživatelská dokumentace se spíše opominutím nedostala do bakalářské práce; ovládání editoru je intuitivní a velkou dokumentaci nepotřebuje, rozhraní je z uživatelského hlediska dostatečně představeno v odeslaném demo paperu.

**Implementační část práce**

lepší    OK    horší    nevyhovuje

Kvalita návrhu    ... architektura, struktury a algoritmy, použité technologie	X			
Kvalita zpracování    ... jmenné konvence, formátování, komentáře, testování	X			
Stabilita implementace	X			

S Františkovou prací jsem byl po celou dobu naší spolupráce velmi spokojen a stejně spokojen jsem i s výsledkem. Velmi kladně hodnotím Františkovu samostatnost při hledání a výběru potřebných komponent (jaká telekonferenční platforma, jaký online editor umožňující současnou práci více uživatelů, ad.) i při jejich zapojování. Architektura je relativně flexibilní a komponenty by nebylo obtížné zaměňovat. Stabilita byla prověřena testem se skutečnými uživateli.

Jako prototyp Františkův Minuteman jasně ověřil praktickou použitelnost konceptu a nástroj mohou technicky zdatní lidé okamžitě reálně nasadit. Vzhledem k atraktivnosti tématu této práce (jak František zmiňuje v sekci 2.2) by bylo velmi vhodné připravit následně takovou podobu aplikace (např. hostované služby), aby ji mohli využít i laici.

**Celkové hodnocení**    Výborně

**Práci navrhuji na zvláštní ocenění**    Ano

Datum

Podpis