

## Správa z workshopu *Automatická transkripcia historických rukopisných a tlačенých dokumentov*

DOI: <https://doi.org/10.24040/ahn.2023.26.02.129-132>

Možnosti a schopnosti digitálnych technológií pomaly, ale isto prenikajú aj do oblasti humanitných vied, konkrétne do oblasti práce s historickými prameňmi, ich spracovávaní, digitalizácie či sprístupňovania. Práve v tomto duchu sa nieslo ďalšie podujatie (workshop) projektu *SKRIPTOR – Inovatívne sprístupnenie písomného dedičstva Slovenska prostredníctvom systému automatickej transkripcie historických rukopisov*, ktorý je podporený Agentúrou pre podporu vedy a výskumu (APVV-19-0456). Workshop sa uskutočnil v dňoch 7. 11. – 8. 11. 2023 v priestoroch Univerzitnej knižnice Univerzity Mateja Bela v Banskej Bystrici, priamo vo vzdelávacom centre Mateja Bela. Okrem nej na podujatí participovala aj Štátna vedecká knižnica v Banskej Bystrici a Filozofická fakulta Univerzity Mateja Bela (Katedra histórie).

Vo workshope môžeme vidieť priamu kontinuitu s predchádzajúcimi podujatiami, ktoré projekt zastrešoval. Z nich asi najvýznamnejšou bola konferencia *Digital humanities – nástroje sprístupňovania historického dedičstva*, ktorá prebehla v októbri 2022. Rovnako ako v jej prípade, aj v prípade workshopu ide o šírenie ideí práce s digitálnymi technológiami v oblasti narábania s písomným historickým dedičstvom. Ako vyplýva z názvu workshopu, orientoval sa na spracovávanie rukopisov a tlačéných dokumentov, a to prostredníctvom platformy *Transkribus*<sup>1</sup>, ktorá priamo ponúka možnosť automatickej transkripcie obsahu dokumentu. Táto schopnosť umelej inteligencie má obrovský potenciál najmä v rámci inštitúcií disponujúcich množstvom dokumentov. Tie je potrebné postupne spracovávať aj do digitálnej podoby, a to nielen vytváraním kópií prostredníctvom fotografovania či skenovania. *Transkribus* totiž disponuje funkcionalitami, ktoré samotnej digitalizácii a transkripcii dávajú značnú pridanú hodnotu. Aj jeden z nosných cieľov projektu hovorí o využití nástroja *Transkribus* na slovacikálne historické rukopisné a tlačéné dokumenty a overení jeho uplatniteľnosti v podmienkach slovenských pamäťových inštitúcií.<sup>2</sup>

---

<sup>1</sup> KATUŠČÁK, D. – NAGY, I. et. al.: Automatická transkripcia historických dokumentov – metodická príručka na prácu s platformou Transkribus. In: *rUMBa – repozitár UMB* [online], 2023, [cit. 28. november 2023]. Dostupné na internete: <https://repo.umb.sk/handle/123456789/412>.

<sup>2</sup> KATUŠČÁK, D. – NAGY, I. et. al.: *Automatická transkripcia historických dokumentov – metodická príručka na prácu s platformou Transkribus*. Banská Bystrica, 2023, s. 5.

Práve z toho dôvodu boli ohľadom účasti na workshope oslovené najmä archívne pracoviská, múzeá, knižnice a výskumné pracoviská. Napokon sa podujatia zúčastnilo 11 záujemcov.

V rámci workshopu bolo potrebné, aby každý z účastníkov bol vybavený notebookom s nainštalovanou platformou *Transkribus*. Vďaka tomu v priebehu zoznamovania sa s jeho funkcionalitami mali účastníci možnosť si sami vyskúšať prácu so softvérom priamo na organizátormi vopred pripravených dokumentoch. Tie boli zhromaždené v zložke určenej na potreby workshopu. Školení záujemcovia mali k dispozícii rôzne typy digitalizovaných dokumentov, čím si mohli osvojiť špecifiká práce s platformou v rámci rôznych druhov písomností. Práve to bolo cieľom workshopu. Dostať *Transkribus* do povedomia pracovníkov pamäťových inštitúcií, zdôrazniť jeho schopnosti, špecifiká práce s ním, nadobudnúť určitú elementárnu orientáciu v rámci platformy a spoločne sa dostať k výsledkom, ktoré môže ponúknuť. Účastníci mali k dispozícii už vyššie citovanú metodickú príručku, kde sú podrobne spracované jednotlivé etapy práce v rámci aplikácie.

Po prezentácii účastníkov a technickej príprave vedúci projektu doc. Mgr. Imrich Nagy, PhD. privítal všetkých zúčastnených a uviedol ich do témy workshopu. Rovnako predstavil členov projektu a výsledky svojej práce. Následne workshop pokračoval svojou školiacou časťou.

Prof. PhDr. Dušan Katuščák, PhD. prezentoval úvod do prostredia *Transkribus* expert klient a jeho funkcionality. Kládol dôraz na nepretržitý vývoj v rámci zdokonaľovania konkrétnych technických zručností platformy a oboznámil prítomných o jej vývoji. Taktiež účastníkov zoznámil so základnými funkcionalitami aplikácie a ich významom, čo spájal s ich využitím v rámci jednotlivých etáp práce.

Nasledoval ďalší blok workshopu, ktorý viedol doc. PhDr. Pavol Maliniak, PhD. Venoval sa príprave dokumentu pred samotnou automatickou transkripciou. Kládol dôraz najmä na prácu so zariadením *ScanTent*<sup>3</sup> a aplikáciou *DocScan*<sup>4</sup>. V rámci tohto bloku zdôraznil potrebu jednotnej metodiky označovania digitalizátov. Vzhľadom na to, že priamo na podujatí bol prítomný *ScanTent*, účastníci si mohli vyskúšať prácu s ním, osvojiť si jeho možnosti či úskalia. Rovnako si do svojich mobilných telefónov nainštalovali *DocScan*, čím si mohli sami nasnímať dokument a presunúť ho do *Transkribu*. Tento postup predstavil PhDr. Oto Tomeček PhD., ktorý sa tiež venoval problematike kvalitného

<sup>3</sup> Zariadenie, ktoré slúži na snímanie dokumentov, s ktorými chceme pracovať v rámci *Transkribu* (viď metodická príručka).

<sup>4</sup> Aplikácia, ktorá priamo súvisí s prácou so *ScanTentom*. Zabezpečuje lepšiu kvalitu snímok a ich plynulý presun do *Transkribu* (viď metodická príručka).

rozlíšenia jednotlivých snímok a jeho otáznemu vplyvu na samotnú chybovosť v rámci transkripcie.

Po úspešnom prenose dokumentov do *Transkribu* obvykle nasleduje segmentácia jednotlivých snímok, čo znamená ich prípravu na fázu automatickej transkripcie. Týmto prítomných previedli Mgr. Michaela Mikušková a Mgr. Lucia Nižníková. Zdôrazňovali dôležitosť procesu segmentácie pre transkripciu, prípadne tvorbu vlastného modelu, ktorý by bolo možné na transkripciu použiť. Taktiež sa venovali konkrétnym príkladom segmentácie v závislosti od typu dokumentu a jeho vonkajšieho charakteru. Ako príklad použili Hurbanove listy (rukopis), Slovenské národné noviny (tlač), Csákósov katalóg korešpondencie Koháryovcov (rukopis formou tabuliek), Orbis pictus (tlač formou tabuliek). Každý jeden krok si spoločne s nimi mohli prechádzať aj účastníci workshopu, keďže mali k dispozícii ukážky dokumentov vo svojich *Transkribus* kontách.

Druhý deň workshopu začala svojim vystúpením Mgr. Alica Kurhajcová, PhD., ktorá prezentovala spôsob tvorby vlastného modelu na automatickú segmentáciu. Ten sa spolu s ňou pokúšali vytvoriť v rámci vlastných kont aj účastníci workshopu (na príklade Hurbanových listov). Zúčastnení si taktiež osvojili proces manuálnej transkripcie jednotlivých snímok dokumentu, ktorá je potrebná pri tvorbe nového modelu. Prednášajúca zdôraznila potrebu jednotnej metodiky práce pri tvorbe modelu a spoločne s členmi workshopu zhodnotili úspešnosť, prípadne neúspešnosť svojich novovzniknutých modelov na automatickú transkripciu. A práve tou sa v ďalšom vystúpení zaoberal doc. Mgr. Imrich Nagy, PhD. Prezentoval postup pri spúšťaní automatickej transkripcie, jej úskalí a možné výsledky. Prítomní mali opäť možnosť si celý tento proces prakticky osvojiť a dospieť k vlastným záverom po nimi spustenej transkripcii dokumentov, ktoré mali vopred k dispozícii.

Ďalej pokračoval blok venujúci sa možnostiam práce s textom po ukončení automatickej transkripcie, ktorý viedla Mgr. Mária Bôbová, PhD. Zamerala sa na možnosti práce s metadátami, ktoré *Transkribus* ponúka, a v rámci nich na využívanie tzv. tagov. Charakterizovala jednotlivé druhy metadát a ich využitie. Venovala sa možnostiam vkladania poznámok v rámci transkribovaného textu. V neposlednom rade tu boli načrtnuté aj možnosti exportu formálnych výstupov automatickej transkripcie a ich ďalšie využitie, čo je konečný cieľ práce v *Transkribe*. Všetky tieto funkcionality si účastníci opäť prakticky osvojili.

Mgr. Patrik Kunec, PhD. zúčastneným predniesol základy práce v *Transkribe* prostredníctvom *Transkribus web app* (tiež aj *Transkribus*

*Lite*), čo je bezplatný<sup>5</sup> webový ekvivalent k *Transkribus Expert Client*. Prítomných previedol jednotlivými časťami tohto webového rozhrania, oboznámil ich s nahrávaním a vytváraním zbierok dokumentov, segmentáciou, automatickou transkripciou či tvorbou vlastného modelu v porovnaní s prácou v *Transkribus Expert Client*.

Ako po každom bloku, aj na záver workshopu bol priestor na diskusiu medzi účastníkmi. Zo strany organizátorov bola ponúknutá možnosť ďalších konzultácií ohľadom práce v *Transkribe* v prípade záujmu o využívanie jeho služieb. Docent Nagy vo svojom záverečnom príhovore poďakoval všetkým zúčastneným a vyzval ich, aby šírili myšlienku automatickej transkripcie a využívania funkcií *Transkribu* ďalej. Taktiež bol načrtnutý úmysel organizovať podujatie takéhoto charakteru aj v budúcnosti. Prípadných ďalších záujemcov, ktorých by takáto forma práce s dokumentmi mohla osloviť, už teraz srdečne pozývame a tešíme sa na ich účasť.

*Matej Smida*

---

<sup>5</sup> Samotná funkcionálnosť transkripcie dokumentu je spoplatnená, rovnako ako pri *Transkribus Expert Client*. Ostatné funkcie sú bezplatné.