
Centrum Malach zažívá zatěžkávací zkoušku. Čeká ho pět měsíců testovacího provozu

<http://tarantula.ruk.cuni.cz/IFORUM-7848.html>

Centrum Malach zažívá zatěžkávací zkoušku. Čeká ho pět měsíců testovacího provozu

Projekt Malach, jehož původním cílem bylo otestovat možnosti automatické indexace multimediálního archivu videonahrávek pamětníků holokaustu, zahájil zkušební provoz Centra Malach. Tento přístupový bod k digitálnímu archivu USC Shoah Foundation se nachází v knihovně fakulty na Malostranském náměstí 25. K dispozici je nejen všem zájemcům z Univerzity Karlovy, ale i dalším odborníkům.

České a slovenské nahrávky, které jsou k dispozici přímo na fakulním serveru, měl na starosti tým z Ústavu formální a aplikované lingvistiky MFF UK. V jaké fázi se nyní projekt Malach nachází, jsme se zeptali jednoho z jeho tvůrců prof. RNDr. Jana Hajiče, Dr., a zástupce Shoah Foundation Institute for Visual History and Education Martina Šmoka.

Od začátku září běží Centrum Malach ve zkušebním režimu. K čemu má sloužit těchto pět měsíců, než bude systém spuštěn v ostré verzi?

M. Šmok: V současné době probíhá zkušební testovací provoz, k oficiálnímu otevření dojde v lednu. Účelem testovací fáze je zjistit, zda vše funguje po technické stránce. Tento čas je také ideální příležitostí pro pedagogy, kteří si mohou přijít vyzkoušet, zda by se jim práce s databází nehodila jako integrální součást jejich výuky. Uživatelé si možná neuvědomují, že archiv neslouží jen k vyhledávání informací, ale i samotná práce s digitálním archivem v reálném čase může být vzdělávacím procesem. Pedagogové tedy mohou přijít se skupinkou studentů a nechat je s databází podle pokynů pracovat.

Komu je archiv přístupný?

M. Šmok: Přijít může kdokoliv. Zatím jsme informaci ale předávali pouze lidem participujícím na projektu nebo těm, u nichž jsme předpokládali, že je tato problematika zajímavá. Třeba i středoškolští učitelé a studenti mohou být další cílovou skupinou. U nich by bylo asi potřeba nejprve udělat kratší seminář o tom, jak s informacemi v archivu zacházet. Měli by si uvědomit, že jde o čistě osobní vzpomínky natáčené po šedesáti letech a že je potřeba pracovat i s jinými zdroji, než jsou pouhá interview.

Jaké další využití má právě zpřístupněný archiv?

M. Šmok: Využití výzkumné. To je otevřeno komukoliv a nemusí se jednat jen o obor historický. Uvědomme si, že jde o soubor 52 000 osobních vzpomínek, dá se tedy jít po fenoménech psychologických, sociologických, etnologických, lékařských, politologických apod.

J. Hajič: Zajímavé je také sledovat, jak mluví lidé z různých míst jedné země či jak hovoří Češi anglicky, tedy ne svým mateřským jazykem. Což může být přínosné pro lingvisty a srovnávací jazykovědce.

M. Šmok: Pro některé jazyky může být archiv největší místní databází daného jazyku. Více mluveného jidiš ve všech regionálních verzích asi nebude nikde jinde v Čechách přístupno.

Jak to bude fungovat prakticky? Musí se návštěvníci předem objednat?

J. Hajič: Podmínky necháme na inženýrce Drahokoupilové, která má centrum na starosti. Po dobu zkušebního provozu asi bude lepší se předem nahlásit. Až budeme fungovat v ostrém provozu, předpokládám, že v určité dny bude možné přijít bez objednání a vždy tu bude volný počítač. Větší skupiny se budou muset z praktických důvodů objednat, aby se tu nesešly dvě třídy najednou.



"Na místní paměti jsou umístěna všechna interview v češtině a slovenštině, to znamená, že jsou uživatelům ihned k dispozici. Vše ostatní se stahuje z centrální databáze v Los Angeles přes Internet2," popisuje princip stahování dat Martin Šmok

Jakou formou budou moci uživatelé s výstupy pracovat?

M. Šmok: Zatím je to jednoznačně on-line digitální archiv k využití na místě. Každý badatel získá své přístupové heslo a bude si moci výsledky svého hledání ukládat a sledovat historii vyhledávání. Pokud například nestihne nějaké interview projít, uloží si ho a vrátí se k němu příště. Pro ty, kteří budou potřebovat další výstupy, je možné vyjednávat s Los Angeles o licenčních ujednáních.

Pokud tedy přijede pedagog z Moravy, který by chtěl ve výuce využít vzpomínky rodáků z dané lokality, tak neodejde s prázdnou?

M. Šmok: Když přijede středoškolský učitel ze Znojma, že by chtěl sestavit hodinu o událostech ve Znojmě, tak myslím, že jsme velice rychle schopni licenční ujednání zařídit a vyrobit kopii příslušného interview. Rozhodovací pravomoci pro licenční podmínky pro Českou republiku má ale opravdu jen Los Angeles.

Jak jsou vlastně data zálohovaná, aby nedošlo třeba při živelných katastrofách k jejich zničení?

J. Hajič: K tomu by opravdu nemělo dojít. Originály jsou uloženy ve dvou sejfech ve Spojených státech. A právě z důvodu přírodních katastrof je jedna kopie na východním pobřeží a druhá na západním.



Prof. RNDr. Jan Hajič, Dr., spoluvůrce projektu Malach, zůstává v klidu. Data by neměla zničit ani případná přírodní katastrofa.

Podívejme se na systém podrobněji. Co všechno umí?

M. Šmok: Vyhledávat lze principem „googlovým“, kdy do pole napíšeme kombinaci libovolných slov a systém začne prohledávat klíčová slova na jakoukoliv kombinaci těchto slov. To je praktické, pokud znáte kódové číslo nějakého interview, jinak dostanete obrovské množství výstupů. Vyhledávání můžete dále specifikovat podle pohlaví, druhu pamětníka, místa a země narození, data narození a řeči nahrávky.

Unikátní je vyhledávání podle klíčových slov. Proklikávat se můžeme kompletním tezaurem nebo si můžeme výraz přesně najít a přiřadit jej do vyhledávače. Pokud si zadám například slovo Terezín, vyjede mi 12 klíčových slov a jejich definice, to vše doplněné o číslo, v kolika interview je slovo použito. Bohužel hodně záleželo na tom, kdo prováděl indexaci nahrávky. Někdo použije slovo tábor a jiný ghetto. Ten, kdo bude pracovat s databází, si musí uvědomit, že to zpracovával také jen člověk, který mohl dělat chyby. Stačí ale s databází chvíli pracovat, abyste si uvědomili, jak systém klíčových slov funguje.

Podle klíčových slov si vyberu nahrávku, která mě zajímá. Co všechno se dozvím o člověku, který v ní hovoří?

M. Šmok: Kromě údajů o místě a datu narození je to například jeho náboženská identita, kde bylo interview natočeno a v bodech, co všechno dotyčný prožil. V rámci interview si můžeme dále prohlížet seznam všech klíčových slov či osob zde zmíněných.

Projekt už je přístupný veřejnosti. Lze tedy systém pokládat za uzavřený?

M. Šmok: Systém se stále vyvíjí. Nedávno byl zprovozněn nový interface, vymýšlí se nové doplňování metadat, připravuje se natáčení svědectví o dalších genocidách, které budou včleněny do tohoto archivu. Prozradit už můžeme, že se rozjíždí například natáčení pamětníků genocidy ve Rwandě.

Více o vzniku  [projektu Malach](#)
(Lucie Kettnerová)