

## Cyber-Search – rekonfigurovatelná platforma se Single Page Interface frontendem

*Bc. Ondřej Smola, Ing. Jiří Jeníček, Ph. D.*

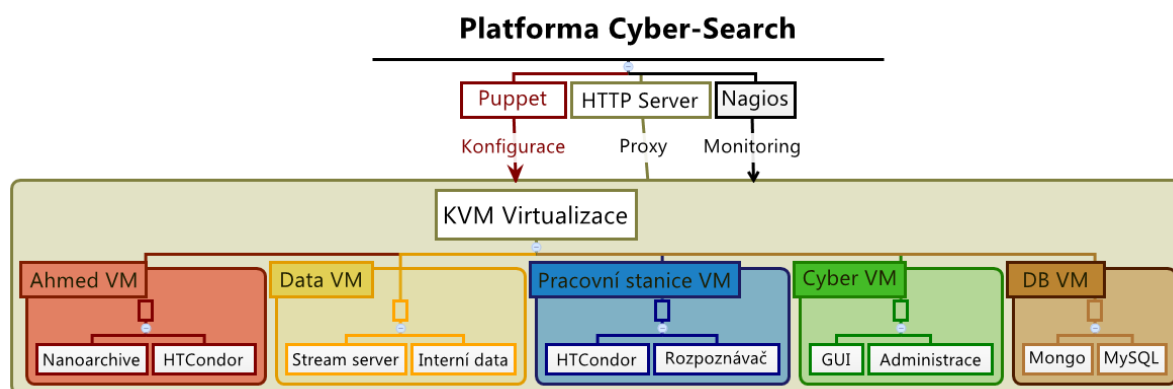
### Abstrakt

Práce se zabývá návrhem a realizací automatizovaného a soběstačného řešení pro systém umožňující offline rozpoznávání mluvené řeči v audiovizuálních materiálech, vyvíjeného na Ústavu informačních technologií a elektroniky Technické Univerzity v Liberci. Výsledkem práce je systém umožňující plně automatizovanou instalaci a konfiguraci této platformy na vlastním hardwaru. Platforma umožňuje automatizované zpracování audio a video souborů a obsahuje veškerou infrastrukturu, potřebnou pro ukládání a prezentaci dat, společně s monitoringem a automatickou údržbou řešení.

### Úvod

Cílem práce je vytvořit plně automatizovaný a centralizovaně spravovaný systém, který využije stávající řešení vyvíjená na Ústavu ITE a vytvoří z nich plně soběstačný a přenositelný celek umožňující zpracování, rozpoznání a indexování nahrávek lidské řeči. Pro prezentaci dat bude vytvořena moderní Single Page Interface webová aplikace s administračním rozhraním. Systém bude umožňovat vzdálenou konfiguraci a dávkové zpracování většího množství dat. Systém musí být navržen tak, aby mohl běžet v libovolném množství instancí a umožnit škálovat dle potřeb od nasazení pro jednotlivé uživatele až po služby s velkým počtem uživatelů. Dále musí být možné centralizovaně upravovat řešení tak, aby umožňovalo pružně reagovat na změny zátěže a v případě potřeby jednoduše přidávat další výpočetní kapacitu.

### Návrh a implementace



Obr. 1: Návrh platformy Cyber-Search

Celý systém se skládá z centrální jednotky spravující platformy Cyber-Search. Všechny platformy Cyber-Search jsou realizovány na virtualizačním řešení KVM a v základní konfiguraci jsou tvořeny 5 virtuálními stroji. Platforma je rozdělena na databázový stroj, datový stroj se streaming serverem a síťovým úložištěm, stroj pro indexaci přepisů a úkolování pracovních stanic a stroj, na kterém běží hlavní Java EE aplikace s uživatelským rozhraním. Pracovní stanice jsou realizovány jako virtuální stroje, sloužící pro výpočetně náročné operace, zastoupené především převodem videa a procesem rozpoznání mluvené řeči.

## Výsledky a diskuze

Výsledkem práce je systém umožňující plně automatizovanou instalaci a konfiguraci platformy na vlastním hardwaru. Pro proces instalace je nutné pouze předání licencí a krátký popis konfigurace. Systém automaticky generuje a provádí veškeré konfigurace, potřebné pro vytvoření platformy, spuštění aplikací, jejich monitoring a automatizovanou správu. Pro přístup a ovládání platformy je implementována Java EE aplikace. Ta je automaticky sestavena a umístěna na aplikační server s pomocí nástroje Maven a serverové aplikace Jenkins. Prezentační vrstva řešení je realizována jako Single Page Interface webová aplikace, vytvořená na základě vlastního aplikačního rámce, vyvinutého během diplomového projektu.

Mezi hlavní výsledky této práce patří kromě samotné realizace výše popsané platformy především popis postupu automatizace komplexního řešení, vyžadujícího rozsáhlou integraci mnoha programů a služeb. Součástí práce je také popis realizace vybraných problematických částí, jako například automatizace operačního systému Windows 7, klasifikace a autentizace klientů během automatizovaného instalačního procesu a diskuze škálovatelnosti platformy.

## Závěr

Vytvořením uceleného systému vznikla platforma, která značnou měrou snížila komplexnost správy rozpoznávače a přidružených systémů a zároveň jim umožnila mnohem větší možnosti vzájemné spolupráce. Tím bylo splněno a zároveň značně rozšířeno zadání práce.

Systém aktuálně spravuje 3 plně aktivní platformy s 6 instancemi. Kromě testovací a vývojové platformy je k dispozici platforma sloužící pro záznamy z jednání Akademického senátu TUL. Aktivní je dále platforma pro prezentaci záznamů přednášek a prezentaci dat potencionálním zájemcům.

Budoucí vývoj práce je nutné směřovat především na dosažení větší samostatnosti systému. Systém aktuálně dokáže automaticky přidávat a konfigurovat virtuální stroje, ale poté již nedokáže bez asistence správce měnit jejich hardwarovou konfiguraci. Mezi hlavní cíle patří možnost realizovat změnu pouze úpravou konfiguračního souboru na centrální jednotce. Dalším krokem je poté s využitím dat z monitoringu umožnit platformě dynamicky reagovat na změny zátěže.

## Poděkování

Tvorba této práce probíhala na základě grantů *Automatická transkripce a indexace přednášek* (TAČR - ALFA, TA01011142), *Zpřístupnění archivu Českého rozhlasu pro sofistikované vyhledávání* (Ministerstvo kultury, DF11P01OVV013) a *Živé Archivy* (TAČR – ALFA, TA01011204). Vývoj práce probíhal na základě výsledků kolektivu Laboratoře počítačového zpracování řeči ústavu ITE a na základě spolupráce s firmou NEWTON Technologies a.s. .

## Reference

BLAVKA, Karel. Indexace a prohledávání multimédií. Liberec, 2010. Dostupné z: <http://knihovna-opac.tul.cz/authorities/431209>. Diplomová práce. Technická Univerzita v Liberci. Vedoucí práce Ing. Jindřich Žďánský, Ph.D.

KLEPÁČ, Martin. Private IaaS cloud comparison. Praha, 2013. Dostupné z: <http://www.diplomovaprace.cz/33/thesis.pdf>. Bakalářské práce. ČVUT. Vedoucí práce Ing. Tomáš Vondra.

JOHNSON, Rod. Expert One-on-One J2EE Design and Development. ISBN 0764543857.