



KIV/SI - Semestrální práce

Vyhledávání klíčových slov v mono/dialogích

Datum:

20.03.2018

Číslo týmu:

5

Autoři:

Matěj Berka (A16N0005K)

Klára Hlaváčová (A16N0036P)

Martin Váňa (A15N0083P)

Michal Veverka (A16N0064P)

Obsah

Obsah	1
1. Zadání	2
1.1. Téma práce	2
2. Analýza	2
3. Implementace	3
3.1 Progressive web app	3
3.2 Vizualizace	3
3.3 Speech-to-text	3
3.4 Frontend	4
4. Uživatelská dokumentace	5
4.1. Prerekvizity	5
4.2. Sestavení	5
4.2.1. Vývojová verze	5
4.2.2. Release verze	5
4.2. Ovládání	6
4.2.1. Nahrání zvukového souboru	6
4.2.1. Přehrání a vizualizace	6
4.2.1. Vyhledání klíčových slov	7
5. Závěr	8
Reference	9
A Screenshots	10

1. Zadání

- V čtyřčlenných týmech vymyslete zadání integračního projektu - něco, co bude mít "wow efekt"
- Dané řešení implementujte v rámci IBM Bluemix s využitím Node.js a dalších komponent

1.1. Téma práce

Cílem práce je vytvořit demo aplikaci, která dokáže vyhledávat klíčová slova v audio nahrávce.

2. Analýza

2.1 Speech-to-text

Převod mluvené řeči na text je jedním z odvětví zpracování přirozeného jazyka a strojového učení a jako takové má důležité místo v moderní informatice. Cílem zpracování řeči je převod lidského hlasu na psaný text. Své využití nalezne v řadě oblastí, např. ovládání hlasem, převod audio záznamů do textové podoby nebo diktování poznámek. Jedním ze způsobů převodu hlasu na text je použití neuronových sítí, konkrétně rekurentních neuronových sítí (RNN). Těch využívá i služba IBM Cloud Speech-to-text [1].

Služba Speech-to-text je poskytuje jednoduché API pro převod mluvené řeči na psaný text. Služba umožňuje získat kompletní přepis analyzovaného audia, ale také vyhledávání klíčových slov v audio. V takovém případě vrátí informace na jakých místech se hledaná slova v záznamu nachází.

Jelikož se jedná o službu IBM Cloud, je třeba nejprve vytvořit účet IBM Cloud a následně získat přístupové údaje (*credentials*) pro službu Speech-to-text. Ty se poté vkládají do příslušných API volání. Službu je možné z aplikace volat pomocí tří rozhraní: websocket, HTTP REST a asynchronní HTTP volání. Nejjednodušší a nejpřímější je použití rozhraní Websocket.

Rozhraní Websocket je možné volat z několika programovacích jazyků (Node, Java, Javascript), nebo také přímo přes HTTP protokol (například pomocí cURL). Posílaná data jsou ve formátu JSON. Službu je možné volat na straně serveru nebo klienta. Všechny způsoby volání služby však vyžadují autentikační údaje. V některých případech (zejména při server-side volání) jsou použity přímo autentikační údaje služby (*service credentials*). Při volání ze strany klienta (Javascript) je však vložení těchto údajů do kódu bezpečnostním rizikem. Proto je u tohoto způsobu použití služby nutné použít tzv. autentikační token. Ten je možné získat pouze voláním ze strany serveru zavoláním autorizačního API IBM Cloud pro danou službu. Každý takový token má životnost 1 hodinu.

3. Implementace

3.1 Progressive web app

Progresivní webová aplikace [8] je hybridem webové stránky a nativní mobilní aplikace. Na mobilním zařízení se aplikace tváří jako nativní - má svojí ikonku v nabídce aplikací, barevné schéma, spouští se ve vlastním okně, a načte se i bez připojení k internetu. Zároveň aplikaci není třeba instalovat a tudíž je snazší ji aktualizovat.

Základem je definice `manifest.json`, který obsahuje metadata o aplikaci. Pomocí technologie ServiceWorkers je umožněno používat aplikaci bez připojení k internetu.

3.2 Vizualizace

Pro vizualizaci nahrávek se uvažovalo o dvou nástrojích a to: Peaks.js a wavesurfer.js. Zpočátku se pro implementaci vybral Peaks.js, protože nabízel, více možností a čitelnější vizualizaci. Během implementace se však narazilo na několik problémů (Kritickým problémem bylo zpracování Blob objektů.), které nakonec vedly k přechodu na wavesurfer.js, který je funkčně o něco chudší, ale pracuje bez problémů.

3.3 Speech-to-text

Službu Speech-to-text voláme z Javascriptu pomocí rozhraní Websocket a knihovny Speech-to-text for Web Browser [2]. K tomuto volání služby je nutné mít autentikační token, který je možné získat pouze ze strany serveru (např. Node.js).

Vytvořili jsme tedy nejprve jednoduchý Node.js server, který při volání `/getToken` zavolá autentikační službu IBM Cloud a vrátí získaný autentikační token. Tento server jsme nasadili do IBM Cloud. Naše aplikace tedy nejprve zavolá tento server a požádá o autentikační token. Tento token poté použije pro volání služby Speech-to-text.

Analýza mluvené řeči je volána pomocí metody `WatsonSpeech.SpeechToText.recognizeFile`. Tato metoda má několik parametrů, nejdůležitější z nich jsou: audio soubor, nastavení analyzátoru, jazykový model, hledaná slova a mnoho dalších. Metoda vrátí *stream*, nad kterým jsou vytvořeny event handlers 'data' a 'error'. Při vyvolání handleru 'data' dojde k příjmu výsledků analýzy ve formátu JSON. Ten obsahuje přepis audia a informace o nalezených klíčových slovech.

3.4 Frontend

Protože se nejedná o žádnou Enterprise class aplikaci, ale pouze jednoduché demo demonstrující integrovanou službu, tak se pro aplikaci vybrala View knihovna React [3] a UI knihovna Semantic UI [4] a neuvažovalo se o některých pokročilejších nástrojích a knihovnách, které by byli v opačném případě vhodnější (Statické typování, testy atd.). Použitou syntaxí je pak ECMAScript 6, která dovolila efektivnější vývoj, a je zpětně překládaná pomocí transpileru na starší verzi JavaScriptu.

Pro samotné sestavení se používá nástroj (React scripts) vytvořený komunitou Reactu, ten vytváří minifikovanou a optimalizovanou verzi aplikace. Na pozadí přitom používá webpack [7].

4. Uživatelská dokumentace

Speech-search je webová aplikace, dostupná jak pro desktopové, tak i pro mobilní zařízení.

Aplikace lze vyzkoušet na adrese: <https://speech-search-final.firebaseio.com/>.

Vzorová data: https://speech-search-final.firebaseio.com/audio_sample.mp3.

4.1. Prerekvizity

- Prohlížeč Google Chrome, verze ≥ 64
- Vývoj: npm ≥ 5.6 (součástí instalace Node.js)

4.2. Sestavení

4.2.1. Vývojová verze

Spuštění na localhostu (ve složce s aplikací)

1. Instalace závislostí: `npm install`
2. Spuštění developer serveru: `npm start` (nebo `npm run start`)
3. Stránka se sama spustí v prohlížeči

4.2.2. Release verze

Před provedením sestavení je nutné nastavit homepage v souboru package.json:

"homepage": "<http://mywebsite.com/relativepath>",

Např.: "homepage": "<http://localhost/speech-search/build>",

Po sestavení se frontend aplikace nachází ve složce *build*.

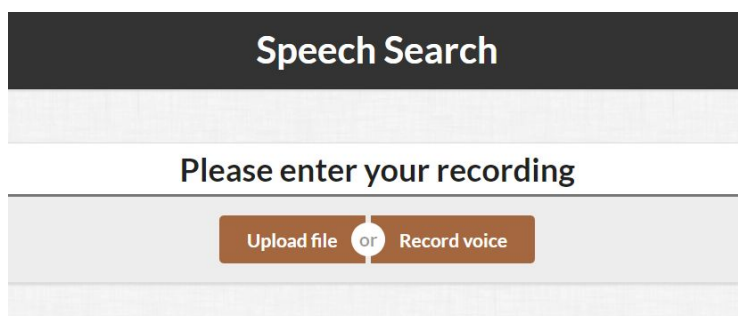
1. Instalace závislostí: `npm install`
2. Sestavení release verze: `npm build` (nebo `npm run build`)

4.2. Ovládání

4.2.1. Nahrání zvukového souboru

Aplikace nabízí dvě možnosti nahrání zvukového souboru: nahrání souboru z disku a nahrání zvuku pomocí mikrofonu. Analyzovaných souborů lze přidat několik.

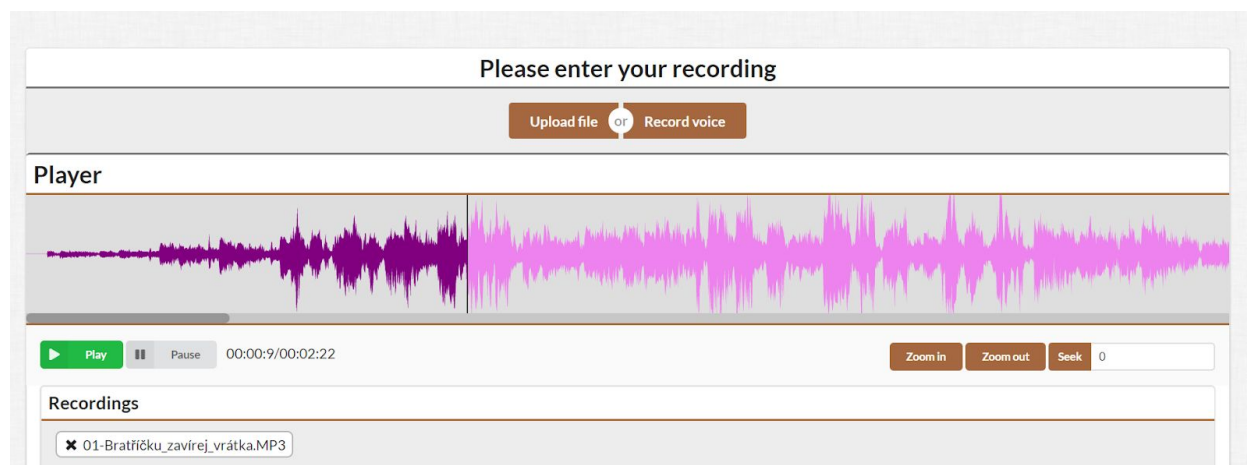
Možný formát zvukového souboru: .wav, .mp3



Obrázek 4.1: Nahrání zvukového souboru

4.2.1. Přehrávání a vizualizace

Po korektním nahrání souboru se zobrazí vizualizace zvukové stopy. Tlačítka pod ní je možné vizualizaci ovládat (přehrát/pozastavit/...)




Obrázek 4.2.: Vizualizace zvukové stopy

4.2.1. Vyhledání klíčových slov

Dostupné jazyky: English (US), English (UK), Japanese, Arabic (MSA, Broadband model), Mandarin, Portuguese (Brazil), Spanish, French (Broadband mode)

Vyhledávaná slova zadáváme do textového pole v oddíle “Keywords”. Tlačítko “Search for keywords” spustí vyhledání slova v nahraném zvukovém souboru. Průběh vyhledávání není nikde zobrazen (vyhledávání nějakou dobu trvá).



The screenshot shows a web interface with two main sections. The top section, titled "Recordings", contains a single item "sample.mp3" with a close button (X). The bottom section, titled "Keywords", contains a single item "some" with a close button (X). Below the keywords, there is a search bar with the placeholder "Keyword", an "Add" button, and a "Search for keywords" button with a magnifying glass icon.

Obrázek 4.3.: Vložení klíčových slov

Po nalezení výskytů slov se zobrazí jejich seznam, informace o slově a ovládací tlačítka pro přehraní a smazání slova. Dále je zobrazen přepis zvukového souboru.

Original keyword	Normalized Text	Confidence	Start Time	End Time	Play	Remove
some	some	92.80000000000001 %	0.18	0.62	<button>Play</button>	<button>Remove</button>
editing	editing	99.8 %	5.42	5.98	<button>Play</button>	<button>Remove</button>

Transcripts

Confidence: 87.5 %

there are some parts that I crippled with because I didn't understand at the time that is in the editing that a film is shaped so I noticed that Steven had shot almost everything I had demanded to know that had to be there but a lot of it ended up cut

Obrázek 4.4.: Seznam výskytů klíčových slov

5. Závěr

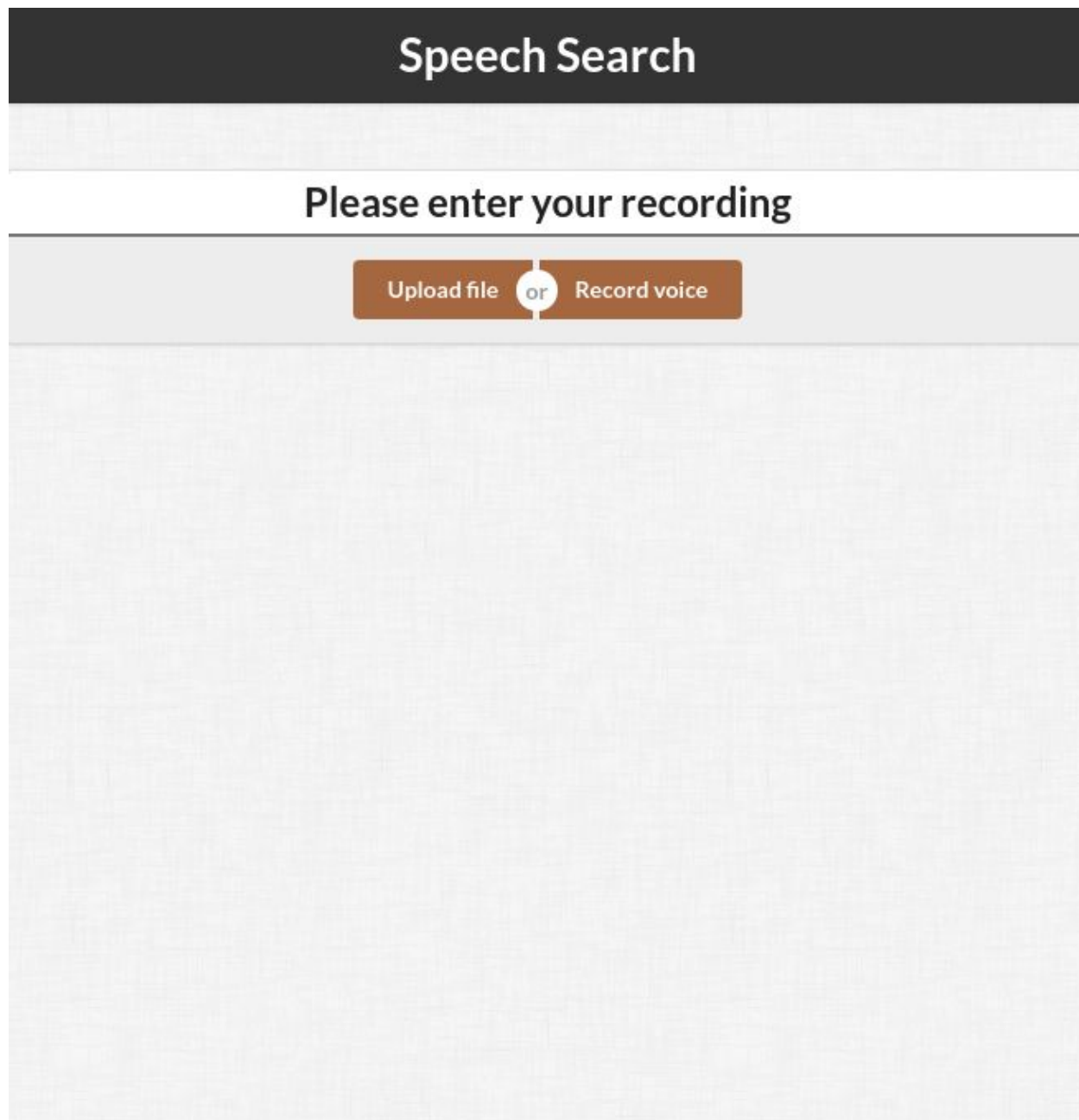
Implementovali jsme demo aplikaci, která dokáže vyhledávat klíčová slova v audio nahrávce, pomocí webových technologií (PWA, React), jednoduchého Node.js serveru a služby z IBM cloudu na převod řeči na text (speech-to-text).

Zadání bylo splněno beze zbytku. Aplikaci založenou na těchto technologiích by šlo použít například pro hledání klíčových slov v audio nahrávkách ze soudní síně.

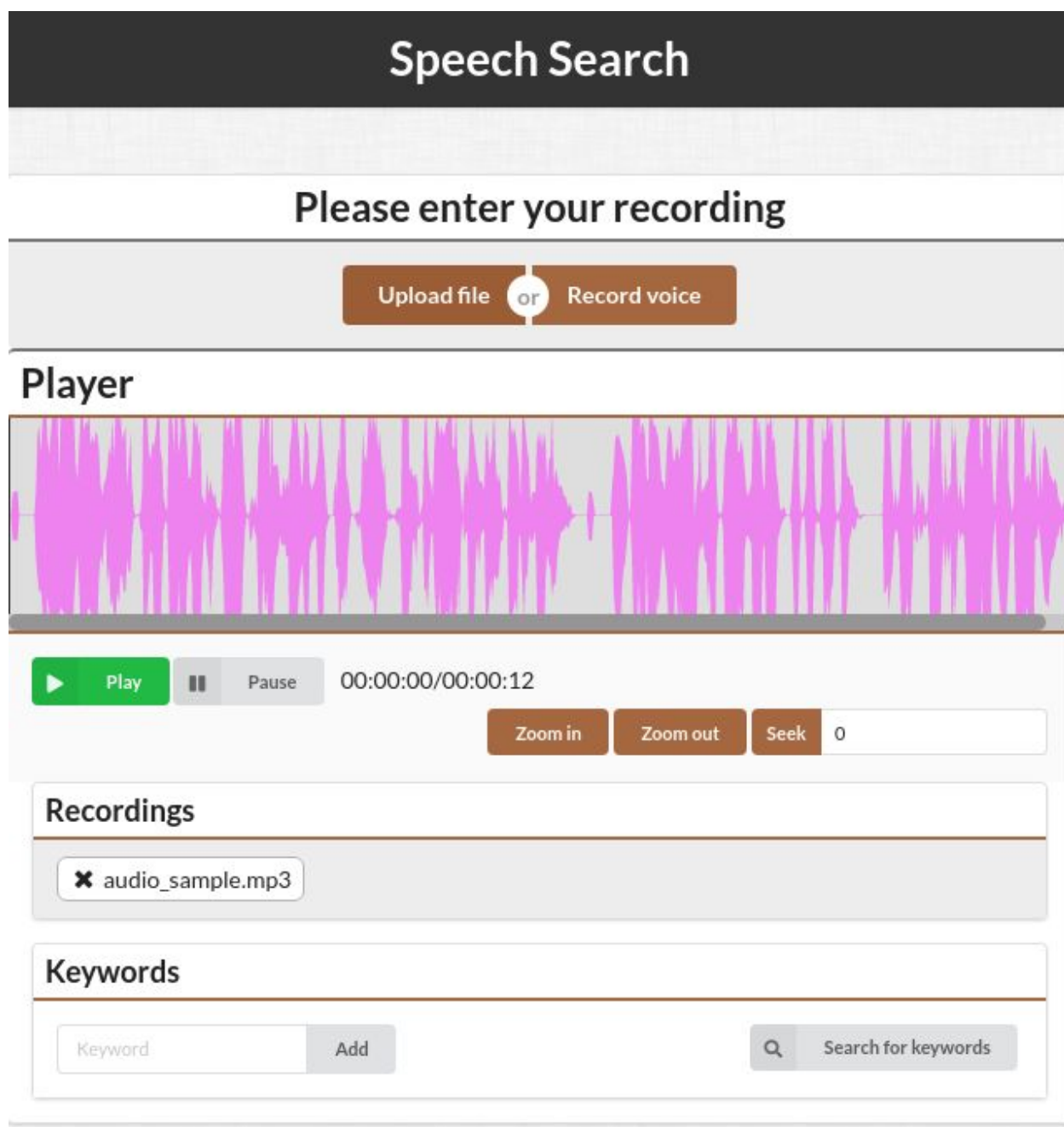
Reference

- [1] - Watson Speech-to-text služba, <https://www.ibm.com/watson/services/speech-to-text/>
- [2] - Knihovna IBM Speech-to-text for Web browser,
<http://watson-developer-cloud.github.io/speech-javascript-sdk/master/>
- [3] - React, <https://reactjs.org/>
- [4] - Semantic UI React, <https://react.semantic-ui.com/introduction>
- [5] - Customizable audio waveform visualization, <https://wavesurfer-js.org/>
- [6] - Peaks.js, <https://github.com/bbc/peaks.js/tree/master>
- [7] - webpack a module bundler <https://webpack.js.org/>
- [8] - Progressive Web Apps <https://developers.google.com/web/progressive-web-apps/>

A Screenshots



Obrázek A.1: Úvodní obrazovka



Obrázek A.2: Po nahrání audio souboru

Speech Search

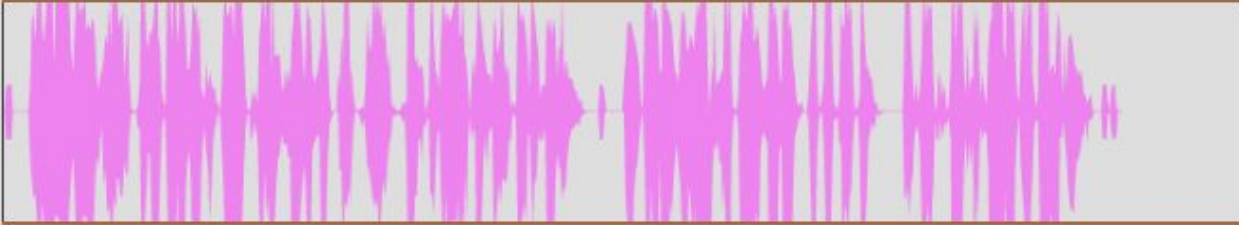
Please enter your recording

Upload file

or

Record voice

Player



▶ Play

⏸ Pause

00:00:00/00:00:12

Zoom in

Zoom out

Seek 0

Recordings

✕ audio_sample.mp3

Keywords

✕ data

Keyword

Add

🔍

Search for keywords

Obrázek A.3: Zadání klíčového slova

Speech Search

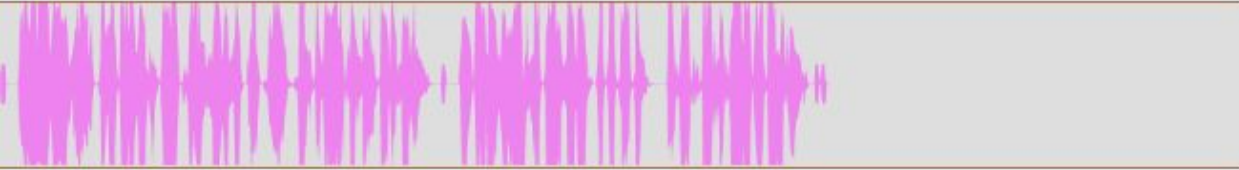
Please enter your recording

Upload file

or

Record voice

Player



Play

Pause

00:00:00/00:00:12

Zoom in

Zoom out

Seek 0

Recordings

✕

audio_sample.mp3

Keywords

✕

data

Keyword

Add

Q

Search for keywords

Original keyword	Normalized Text	Confidence	Start Time	End Time	Play	Remove
data	data	81.2 %	5.3	5.54	<div>Play</div>	<div>Remove</div>
data	data	86.4 %	6.52	6.91	<div>Play</div>	<div>Remove</div>

Transcripts

Confidence: 87.5 %

windows as your is an internet scale cloud services platform hosted in Microsoft data centers your data your network in your business are protected by H. B. security technologies

Obrázek A.4: Finální obrazovka po nalezení výskytu klíčových slov a zobrazení přepisu