VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV INTELIGENTNÍCH SYSTÉMŮ
DEPARTMENT OF INTELLIGENT SYSTEMS

DATABÁZE VYSÍLÁNÍ ČESKOSLOVENSKÉ TELEVIZE

DATABASE OF CZECHOSLOVAKIAN TELEVISION BROADCASTING

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — **VOJTĚCH FIALA**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. JAROSLAV ROZMAN, Ph.D.**
SUPERVISOR

BRNO 2022

Vysoké učení technické v Brně
Fakulta informačních technologií

Ústav inteligentních systémů (UITS) Akademický rok 2021/2022

Zadání bakalářské práce

Student: **Fiala Vojtěch**
Program: Informační technologie
Název: **Databáze vysílání Československé televize**
Database of Czechoslovakian Television Broadcasting
Kategorie: Umělá inteligence

Zadání:

1. Nastudujte OCR program Československé televize v Týdeníku televize. Nastudujte práci s relačními databázemi.
2. Navrhněte a vytvořte program, který automaticky z OCR textu z Týdeníku televize identifikuje části s televizním programem a určí časy vysílání a názvy pořadů.
3. Navrhněte relační databázi, do které se extrahovaný program uloží. Zároveň navrhněte jednoduché webové rozhraní, které umožní program prohlížet a filtrovat různé pořady.
4. Navrženou databázi a web implementujte.
5. Proveďte zhodnocení úspěšnosti automatické tvorby televizního programu.

Literatura:

- J. Adamus, J. Hošek, Program 68, https://www.ceskatelevize.cz/specialy/totostoleti/program-68, [cit. 15.10.2020]

Pro udělení zápočtu za první semestr je požadováno:

- První tři body zadání.

Podrobné závazné pokyny pro vypracování práce viz https://www.fit.vut.cz/study/theses/

Vedoucí práce: **Rozman Jaroslav, Ing., Ph.D.**
Vedoucí ústavu: Hanáček Petr, doc. Dr. Ing.
Datum zadání: 1. listopadu 2021
Datum odevzdání: 11. května 2022
Datum schválení: 3. listopadu 2021

Abstrakt

Tato práce se zabývá vytvořením databáze vysílání Československé televize za pomoci technologie optického rozpoznávání znaků. Cílem práce je vytvořit databázi, která bude přístupná skrze jednoduché webové rozhraní, ve kterém bude možné prohlížet jednotlivé pořady, jejich popisy a mezi pořady vyhledávat. Vyhledávání bude možné i podle osob, co se na pořadech podílely. Jako systém optického rozpoznávání znaků byl použit Tesseract, získané pořady byly nahrány do SQLite databáze a pro webové rozhraní byl použit Flask. Jako zdrojové materiály byly použity naskenované strany Týdeníku Československé televize a zpracovány byly roky 1966 až 1992. Výsledky práce umožňují uživatelům nahlédnout do historie vysílání televize v Československu a mohou sloužit například pro účely historiků.

Abstract

This thesis describes the process of creating a Czechoslovak Television broadcasting database using optical character recognition technology. The aim of the work is to create a database that can be accessed through a simple web interface, in which it will be possible to browse individual shows, their descriptions and to search among the shows. Searching will also be possible by people involved in the shows. Tesseract was used as the optical character recognition engine, obtained shows were entered into a SQLite database and Flask was used for the web interface. Scanned pages of Týdeník Československé televize were used as source materials and the years from 1966 to 1992 were processed. The results of the work give users an insight into the history of television broadcasting in Czechoslovakia and can be used, for example, for the purposes of historians.

Klíčová slova

OCR, Tesseract, relační databáze, televizní vysílání, databáze vysílání, televizní program, Československá televize

Keywords

OCR, Tesseract, relational database, television broadcasting, broadcasting database, television programme, Czechoslovakian television

Citace

FIALA, Vojtěch. *Databáze vysílání Československé televize.* Brno, 2022. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Jaroslav Rozman, Ph.D.

Databáze vysílání Československé televize

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Jaroslava Rozmana, Ph.D. Uvedl jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpal.

.....
Vojtěch Fiala
5. května 2022

Poděkování

Chtěl bych poděkovat svému vedoucímu, Ing. Jaroslavu Rozmanovi, Ph.D., za vedení, konzultace a dobré rady. Dále bych chtěl poděkovat pracovníkům České televize za poskytnutí materiálů pro vypracování této práce. Děkuji také rodině a přátelům za jejich podporu nejen při tvorbě této práce, ale po celou dobu studia.

Obsah

Kapitola 1

Úvod

Televizní programy existují v podstatě od počátků televizního vysílání. Původně byly vydávány pouze v papírové podobě, dnes jsou běžně dostupné i na internetu. V případě historických televizních programů tomu ovšem tak není. Běžný člověk si může chtít připomenout dobu minulou a nebo například v rámci vědecké práce analyzovat vysílané pořady. To však není bez fyzické návštěvy archívu možné. Digitální verze televizního programu by tento problém vyřešila a umožnila uživatelům si historický program prohlížet odkudkoliv, pouze s využitím vlastního počítače.

Tato práce si klade za cíl vyřešit výše zmíněné problémy a vytvořit databázi vysílání Československé televize z let 1966 až 1992. Využity pro tento účel budou naskenované strany tehdejšího *Týdeníku Československé televize*, jehož ukázku lze vidět v příloze B. Z těchto stran bude za použití optického rozpoznávání znaků získán text televizního programu, který bude uložen do databáze. Ta bude zpřístupněna skrz jednoduché webové rozhraní. Toto rozhraní bude umožňovat snadné prohlížení pořadů, mezi kterými bude možné také vyhledávat.

V této práci bude popsán návrh sady nástrojů, které zajistí proces digitalizace, a budou představeny odlišné metody, jak k tomuto úkolu přistoupit. Hned v následující kapitole 2 bude pro uvedení čtenáře do problematiky představena základní funkcionalita systémů pro optické rozpoznávání znaků. Dále bude v kapitole 3 následovat úvod do problematiky relačních databází. Kapitola 4 se poté bude zabývat obecnějším návrhem sady nástrojů pro digitalizaci televizních programů a jejich dodatečnou úpravou, návrhem databáze, do niž budou uloženy zdigitalizované programy, a nakonec také návrhem jednoduchého webového rozhraní pro přístup k nim. Kapitola 5 bude obsahovat již konkrétní popis implementace navržených řešení. V poslední kapitole 6 dojde k vyhodnocení úspěšnosti procesu digitalizace a návrhu dalšího vývoje.

Kapitola 2

Optické rozpoznávání znaků

Tato kapitola se zabývá uvedením čtenáře do základů problematiky optického rozpoznávání znaků. Nejedná se o encyklopedický přehled, ale pouze o soupis relevantních informací a uvedení do problematiky. Jako první budou vysvětleny základy jednotlivých fází procesu zpracování textu s využitím optického rozpoznávání znaků (dále jen OCR – Optical Character Recognition) a následně budou představeny některé z již existujících nástrojů pro zpracování textu za pomoci OCR.

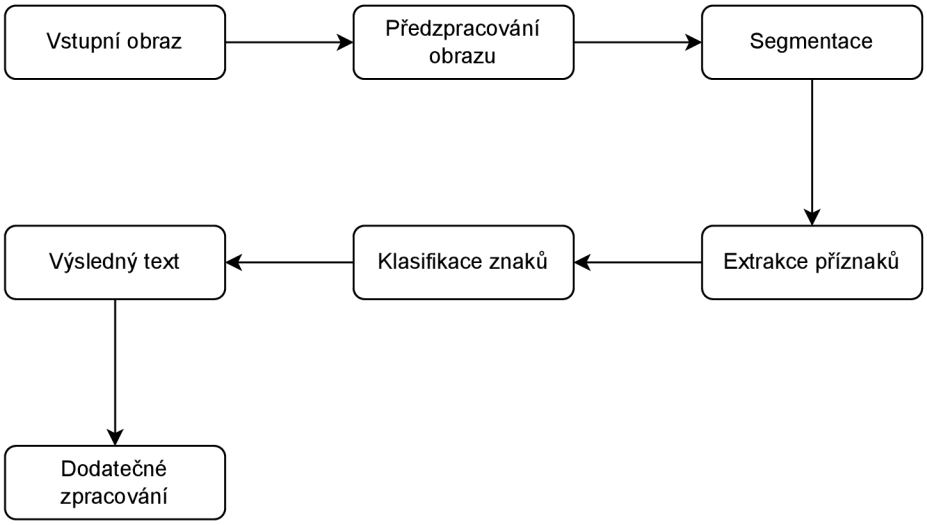
OCR je proces zpracování ručně psaného či tištěného textu počítačem. Tento text je potřeba nejprve převést do digitální podoby, například jeho vyfocením či naskenováním. Systém OCR následně zpracuje poskytnutý obraz a vrátí text, který se na vstupním obrazu nacházel.

V současnosti existuje mnoho systémů OCR, které jsou schopné dosahovat přesnosti až 99 % [4]. Lze je rozdělit na komerční a volně dostupné. Komerční systémy jsou placené s uzavřeným zdrojovým kódem, volně dostupné systémy může pro svoji potřebu používat a upravovat každý. Existující systémy nejsou dokonalé a samotný proces zpracování textu pomocí OCR je výpočetně náročný. V dnešní době jsou pro fungování systémů OCR často využívány neuronové sítě, což zajišťuje oproti klasickým přístupům větší efektivitu [2].

2.1 Zpracování textu

Zpracování textu systémem OCR probíhá v několika fázích. Úkolem každé fáze je poskytnutý vstup připravit pro zpracování další fází tím, že se sníží množství nepotřebných informací. To znamená, že každá fáze pracuje s méně informacemi než fáze předchozí. Výsledkem procesu zpracování je pak získaný text v digitální podobě. Následně je vhodné kvalitu získaného textu navýšit dodatečným zpracováním. Celý proces je graficky znázorněn na obrázku 2.1, který volně vychází z procesu zpracování s využitím OCR tak, jak je popsán v [2], včetně dodatečného zpracování.

Náročnost každé fáze se liší v závislosti na vstupním zdroji textu – například u krátkého textu, jakým může být pouze jedno slovo, bude jeho rozdělení na písmena jednodušší, než by tomu bylo u celé popsané strany.



Obrázek 2.1: Schéma zpracování obrazu za pomoci OCR

2.1.1 Předzpracování

První fází v procesu zpracování textu systémem OCR je předzpracování vstupního obrazu. Pro zajištění nejlepších výsledků je nutné vstupní obraz transformovat do vhodného formátu. Při tom dochází k odstranění zbytečných informací, jako jsou barvy či případný obrazový šum. Nakonec může být vhodné provést další korekce, například v podobě rotace textu do vhodného úhlu. [2]

K základním metodám předzpracování patří prahování, při kterém je barevný vstupní obraz převeden do černé a bílé barvy. Prahovacích metod existuje mnoho, lze je ale rozdělit na dva specifické okruhy – *globální* a *lokální* [3].

Globální prahování

Jedná se o nejjednodušší metodu prahování, kdy je zvolen jeden práh, oproti kterému je následně porovnán jas každého pixelu v obrazu. V případě, že je jas daného pixelu vyšší než hodnota prahu, je barva pixelu nastavena jako bílá. V opačném případě je nastavena na černou. Tento proces je popsán následující rovnicí 2.1 vycházející z [5].

$$bod(x,y) = \begin{cases} 255 & \text{jestli } bod(x,y) > \text{práh} \\ 0 & jinak \end{cases} \tag{2.1}$$

Tuto metodu je vhodné použít v případě, že se sledované objekty, v tomto případě písmena, barevně odlišují od okolí a všechny mají stejnou, nebo alespoň velmi podobnou, barvu a jas [3]. Ukázku výsledku použití globálního prahování je možné vidět na obrázku 2.2.

(a) Původní obraz

(b) Obraz po provedení globálního prahování

Obrázek 2.2: Ukázka efektu globálního prahování

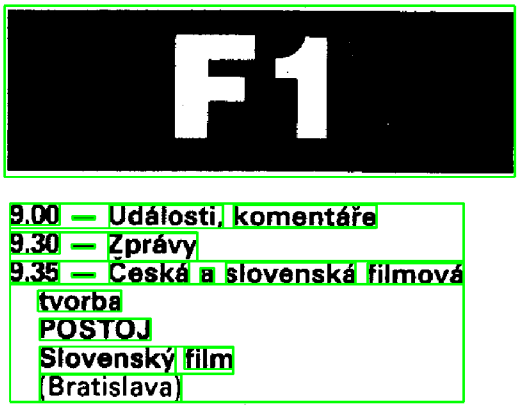
Další variantou globálního prahování je *Otsuho metoda* [3]. V čem se tato metoda odlišuje, je to, že práh je volen automaticky na základě výpočtu. Otsuho metoda prahování je vhodná například v případě, kdy je potřeba zprahovat vícero obrazů, přičemž každý z nich má jiný jas, a manuálně volit globální práh pro použití u každé z nich by bylo časově náročné.

Lokální prahování

Od globálního prahování se lokální liší v tom, že nepoužívá pro celý obraz pouze jeden globální práh. Zatímco v případě globálního prahování jsou jednotlivé pixely porovnávány s předem stanovenou pevnou hodnotou, lokální prahování používá pro každý pixel odlišný práh. Hodnota prahu je získávána na základě analýzy okolních pixelů. Tento přístup je vhodný v situaci, kdy má vstupní obraz odlišný jas v různých částech – to může být zapříčiněno například stíny či grafickou úpravou obrazu. [7]

2.1.2 Segmentace

Výsledkem předchozí fáze je vyčištěný obraz od rušivých jevů. Následující fází v procesu optického rozpoznávání znaků je segmentace. Cílem této etapy je detekovat text a rozdělit jej na jednotlivé bloky v závislosti na rozložení strany. Bloky jsou dále děleny na jednotlivé řádky, přičemž řádky jsou děleny na jednotlivá slova, například za pomoci detekování mezer [2]. Rozdělení textu do bloků, které jsou rozděleny na jednotlivá slova, lze vidět na obrázku 2.3.



Obrázek 2.3: Vstupní obraz segmentovaný na bloky textu obsahující jednotlivá slova

Získaná slova jsou dále rozdělena na jednotlivá písmena za účelem následné klasifikace [6]. Na obrázku 2.4 je možné vidět výsledek procesu segmentace – vstupní text, který byl rozdělen na jednotlivá písmena, jež budou následně systémem OCR dále zpracována.

Obrázek 2.4: Text ze vstupního obrazu, segmentovaný na jednotlivá písmena

Obzvlášť v případě ručně psaného textu se může jednat o velmi problematickou část, neboť nemusí být zachovány řádky ani dostatečné mezery mezi slovy. Jednotlivé znaky se

navzájem také mohou, ale nemusí, dotýkat, což může mít za následek neúspěšnou segmentaci znaků ve slově [12].

Další problematickou částí může být segmentace složitě členěného obrazu. U složitěji komponovaných dokumentů může dojít k nesprávné segmentaci textových částí do bloků, což může vést například k tomu, že text rozdělený vizuálně na sloupce je segmentován nikoliv jako vícero sloupců, ale jako řádky.

2.1.3 Extrakce příznaků

V této fázi jsou z jednotlivých segmentovaných písmen, které byly výsledkem předchozí fáze, extrahovány jejich základní příznaky. Každý znak je má odlišné. Mezi tyto příznaky může patřit například počet čar a oblouků, které písmeno tvoří. To umožňuje jednotlivé znaky na základě jejich příznaků od sebe rozpoznat a v další fázi poté identifikovat. [2]

2.1.4 Klasifikace znaků

Předposlední fází je klasifikace znaků na základě jejich příznaků, které byly získány jako výsledek předchozí fáze. Cílem této fáze je určit již konkrétní znaky, které se nacházely v původním textu. Způsob, jakým je znak klasifikován, závisí na příznacích, které byly ze znaků extrahovány v předchozí fázi. Možných způsobů klasifikace je mnoho. Mezi základní z nich patří například metoda porovnávání šablon. [2]

Porovnávání šablon

Metoda porovnávání šablon patří mezi ty nejjednodušší, co se komplexity týče. Na základě příznaků extrahovaných v rámci minulé fáze dochází k porovnání těchto příznaků s příznaky odpovídajícími šablonám jednotlivých znaků. Toto porovnání může být triviálního charakteru, kdy jsou se šablonami srovnávány pixely, z nichž jsou písmena složena, ale také může být složitější, kdy jsou porovnávány pouze vybrané části, nad kterými mohou být prováděny další operace. [2]

2.1.5 Dodatečné zpracování výsledku

V poslední fázi už je k dispozici načtený text z původního obrazu. Nad ním je prováděno dodatečné zpracování, které se děje za účelem dalšího zvýšení kvality výsledků. Text, jenž byl získán jako výsledek klasifikačních procesů z minulé fáze, nemusel být z podstaty fungování OCR klasifikován správně. Některé z chyb, ke kterým mohlo dojít, jsou popsány již u segmentace. Problémů, které mohly nastat, je ale více. Vstupní obraz nemusel být dostatečné kvality a nebo mohlo dojít k použití nevhodné metody pro předzpracování obrazu. Takto nevhodně zvolená metoda mohla vést k výskytu nežádoucího šumu, který se jejím použitím nepovedlo eliminovat. Dalším problémem, ke kterému mohlo dojít, je nesprávná klasifikace znaku. Mimo to existuje mnoho jiných problémů, které zde není důvod encyklopedicky vyjmenovávat.

Tyto problémy je možné se pokusit vyřešit dodatečně. To platí obzvláště v případě, kdy nesprávně detekováno bylo například pouze jedno písmeno ve slově – čím blíže je výsledek k originálu, tím větší jsou šance, že se jej povede opravit. Pro tyto účely dodatečného zpracování je nejčastěji využíváno slovníku pro zdrojový jazyk, ve kterém je napsán vstupní text. Jednotlivá slova, která jsou systémem načtena, jsou oproti slovníku porovnána. V pří-

padě, že ve slovníku nejsou nalezena, jsou změněna na slova, která se zadané posloupnosti znaků nejvíce podobají. [2]

Nevýhodou této metody je, že je zdlouhavá, neboť porovnání načteného slova oproti všem slovům daného jazyka může být velmi časově náročné. Další riziko v případě používání slovníku může nastat v souvislosti s vlastními jmény. Správně detekované vlastní jméno může být slovníkem chybně opraveno na jiné slovo, protože slovník vlastní jména neobsahuje a nebo jich obsahuje pouze omezené množství.

2.2 Existující nástroje

V podkapitole 2.1 byl popsán obecný proces fungování systému OCR. Protože by však implementace takového systému pokaždé znovu byla náročná a drahá, existuje na trhu mnoho již hotových řešení, která si lze přizpůsobit pro konkrétní potřeby. Každý z těchto nástrojů je vhodný pro odlišné použití pod odlišnou licencí. Mnoho z nich je placených a nebo s uzavřeným zdrojovým kódem – například *Amazon Textract* či *Microsoft Azure Computer Vision*. Tyto komerční systémy OCR nejsou kvůli svým licenčním podmínkám pro tuto práci vhodné. Jak bylo řečeno již v úvodu, existují ale i volně dostupné nástroje, které může pro svoji potřebu volně využívat každý. Některé z nich budou představeny v následujících řádcích.

Tesseract

Tesseract[1] je volně dostupný systém s otevřeným zdrojovým kódem pod licencí Apache 2.0. Původně byl vytvořen už v roce 1984 firmou Hewlett-Packard, v roce 2005 byly jeho zdrojové kódy zveřejněny a v letech 2006-2018 byl vyvíjen firmou Google. Tento systém je dostupný pro OS Windows, OS Linux a macOS. Podporuje více než stovku jazyků včetně češtiny a umožňuje uživatelům systém trénovat na vlastních datech a podporu rozšiřovat. Dále nabízí uživatelům možnost zvolit si způsob automatické segmentace vstupního textu. Jeho současná verze, kterou byla v době psaní této práce verze 5.0, využívá LSTM neuronové sítě (ty jsou využívány pro fungování od verze 4.0), což vedlo oproti původní verzi ke zvýšení přesnosti výsledků.

EasyOCR

EasyOCR[2] nabízí více než 80 podporovaných jazyků, opět včetně češtiny. Stejně jako již zmíněný Tesseract je i EasyOCR volně dostupný systém s otevřeným zdrojovým kódem, dostupný pod licencí Apache 2.0. Oproti Tesseractu se jedná o méně rozsáhlý projekt, za nímž stojí společnost Jaided AI.

OCRopus

OCRopus[3] je dalším volně dostupným systémem pod licencí Apache 2.0. Slouží pro analýzu dokumentů a provedení optického rozpoznávání znaků na nich. Modularita nástroje umožňuje provádět v případě potřeby pouze určité kroky OCR zpracování.

[1] https://github.com/tesseract-ocr/tesseract
[2] https://github.com/JaidedAI/EasyOCR
[3] https://github.com/ocropus/ocropy

Kraken

Kraken[4] je oproti předchozím zmíněným navržen převážně pro práci s historickými dokumenty. Opět se jedná o volně dostupný systém. Dostupný je, stejně jako předchozí zmíněné, pod licencí Apache 2.0. Dostupný je pouze pro Linux a MacOS, Windows není podporován.

Pro využití v této práci byl zvolen systém OCR **Tesseract**. V jeho prospěch hrají kvalitní dokumentace, široká uživatelská základna (a tedy i uživatelská podpora), ale hlavně kvalita rozpoznávání slov, znaků, a schopnost text automaticky segmentovat do bloků, což je pro tuto práci klíčové.

[4] https://kraken.re

Kapitola 3

Relační databáze

V této kapitole budou představeny základní prvky relačních databází, proces normalizace databází a nakonec budou krátce popsány existující databázové systémy. Opět budou představeny pouze informace pro tuto práci relevantní.

Databázi [9] lze charakterizovat jako soubor vzájemně propojených prvků, nad nimiž operuje systém řízení báze dat – speciálně optimalizovaný systém sloužící pro operace nad těmito záznamy. Mezi tyto operace patří například vkládání nových záznamů, aktualizace stávajících záznamů či jejich odstraňování.

Relační databáze [9] jsou jedním z podtypů databází. Jsou založeny na relačním modelu dat a relační algebře. Relace jsou zde reprezentovány například tabulkami databáze a nebo také vztahy mezi těmito tabulkami. V tabulkách relační databáze jsou následně uloženy záznamy, které mají podobu n-tic, do jednotlivých řádků. Pro dotazování nad daty uloženými v relační databázi se obvykle používá jazyk *SQL* [18]. Je navržen tak, aby se více podobal běžnému jazyku než typickému programovacímu jazyku a příkazy v něm jsou většinou intuitivní.

Jednotlivé tabulky jsou složeny z řádků a sloupců. Sloupce tabulky, také označované jako *atributy*, obsahují vždy hodnoty stejného typu. Každý řádek v tabulce musí obsahovat *primární klíč* – jednoznačný identifikátor daného řádku. Primární klíče mohou být tvořeny hodnotami z jednoho sloupce, ale mohou být tvořeny i více sloupci tabulky – v takovém případě tyto sloupce tvoří *složený primární klíč*. V tabulkách je možno sloupec označit jako *cizí klíč*, což značí odkaz na jinou tabulku, kde jsou hodnoty v daném sloupci použity jako klíče primární. Tabulka může obsahovat více unikátních hodnot, které umožňují jednoznačně identifikovat celý řádek – všechny tyto hodnoty jsou označovány jako *kandidátní klíče*, přičemž zmíněný primární klíč je tvořen některým z nich. [9]

3.1 Normalizace dat

Cílem procesu normalizace dat je odstranit z databáze redundantní záznamy. To je prováděno za účelem zvýšení efektivity databázového systému a zjednodušení úprav dat. Bez jakékoliv normalizace by databáze obsahovala položky více než jednou, což by vedlo k neefektivitě celého systému, jelikož by se ukládaly nepotřebné informace. Toto vícenásobné ukládání informací u databází, které nejsou normalizovány, má také za následek problematickou úpravu dat. V případě, že by nastala potřeba nějaký údaj upravit, musel by být upraven na více místech, což je zbytečné jak z hlediska času, tak z hlediska systémových zdrojů.

Běžně používaných normálních forem je 5, případně 6 za podmínky, že je Boyce-Coddova (3,5.) normální forma považována za samostatnou. Nejčastěji se využívá 1., 2. a 3. normální forma. Vyšší normální formy slouží již pouze k zajištění efektivity při používání složených primárních klíčů [9]. Data se nacházejí v n-té normální formě právě tehdy, když splňují konkrétní podmínky, které daná normální forma vyžaduje, a zároveň se nacházejí i ve všech nižších normálních formách.

První normální forma

V první normální formě jsou data právě tehdy, když žádný ze sloupců tabulek neobsahuje víceatributová data – všechny hodnoty jsou *atomické*. [9]

Druhá normální forma

Databáze splňuje podmínky druhé normální formy tehdy, když jsou data zároveň v první normální formě a navíc jsou všechny neklíčové atributy *funkčně závislé*[1] na celém primárním klíči, tj. žádný z neklíčových atributů nesmí funkčně záviset pouze na části případného složeného primárního klíče. Lze konstatovat, že každá tabulka v databázi, která splňuje podmínky první normální formy a zároveň neobsahuje složený primární klíč, je automaticky v druhé normální formě. [9]

Třetí normální forma

Podmínky třetí normální formy splňuje databáze právě ve chvíli, kdy splňuje podmínky druhé normální formy a zároveň se v ní nevyskytují *tranzitivní závislosti*[2], tj. žádný neklíčový atribut není funkčně závislý na žádném jiném neklíčovém atributu. Všechny neklíčové atributy musí být závislé pouze na primárním klíči. [9]

Boyce-Coddova normální forma

V Boyce-Coddově normální formě (také označované jako rozšířená třetí normální forma, nebo také 3,5. normální forma) se databáze nachází v případě, že splňuje podmínky třetí normální formy a zároveň platí, že neexistuje žádný atribut, který by nebyl primárním nebo kandidátním klíčem, na němž by závisel jiný atribut. To lze chápat tak, že žádný atribut, který není primárním klíčem, nesmí určovat žádný jiný atribut – ani část primárního klíče. [9]

Další normální formy

Vyšší normální formy, částečně včetně již zmíněné Boyce-Coddovy normální formy, jsou pouze dodatečnými, neboť v praxi stačí 3. normální forma pro vyřešení 90 % problémů, které mohou nastat [9]. Proto zde tyto formy již nebudou podrobněji rozebírány.

[1] Funkční závislost znamená, že se konkrétní hodnota atributu vždy odvíjí od hodnoty jiného atributu

[2] Tranzitivní závislost znamená, že je atribut funkčně závislý na jiném, neklíčovém atributu, který závisí na primárním klíči

3.2 Existující relační databázové systémy

Existujících řešení v podobě systémů implementujících relační databáze je mnoho. Jsou k dispozici jak komerční, tak i volně dostupné databázové systémy. Mezi nejpoužívanější z nich patří mimo jiné **MySQL**, **MariaDB** či **SQLite** [13]. Představením těchto nejpoužívanějších volně dostupných řešení se zabývají následující řádky.

MySQL

MySQL [10] je nejpopulárnější volně dostupný databázový systém s otevřeným zdrojovým kódem, dostupný pod licencí GNU/GPL. Je vyvíjen firmou Oracle Corporation. Mezi hlavní cíle vývoje MySQL patří rychlost, spolehlivost, jednoduchost použití a schopnost pracovat s velkým množstvím dat.

MariaDB

Mezi další nejpoužívanější databázové systémy s otevřeným zdrojovým kódem patří MariaDB [8]. Je vyvíjena originálními autory MySQL. Původně měla sloužit jako lepší náhrada za MySQL, z čehož plyne, že je s MySQL kompatibilní. Systém je dostupný pod licencí GNU/GPLv2 a mezi významné uživatele tohoto systému patří mimo jiné Google, Wikipedie či WordPress.

SQLite

Oproti výše zmíněným databázovým systémům je SQLite [14] mírně odlišný. Stejně jako výše zmíněné má otevřený zdrojový kód, jenž v tomto případě ale není chráněn žádnou licencí – jedná se o volné dílo. V čem se hlavně odlišuje, je to, že nemá, narozdíl od většiny ostatních databázových systémů, oddělený server. Zápis a čtení probíhají přímo do, respektive z, běžných souborů na disku. Databáze vytvořené systémem SQLite jsou také multiplatformní. SQLite si klade za cíl být rychlý, kompaktní a spolehlivý.

Pro použití v této práci byl zvolen systém **SQLite**. Jeho hlavními výhodami, které k jeho výběru vedly, jsou velmi jednoduché zprovoznění, používání a také schopnost fungovat na více platformách.

Kapitola 4

Návrh řešení

V rámci této kapitoly bude provedena analýza požadavků řešení a dojde k rozboru formátu Týdeníku ČS televize. Na základě tohoto rozboru dále dojde k návrhu sady nástrojů, které zajistí automatickou extrakci vysílaných pořadů z digitálních skenů Týdeníku ČS televize. Následovat bude návrh relační databáze, do níž budou extrahované pořady ukládány. Závěrem bude v této kapitole představen návrh jednoduchého webového rozhraní, skrz které bude možné k databázi vysílání přistupovat.

4.1 Analýza požadavků a základní návrh jejich řešení

Cílem této práce je navrhnout a implementovat relační databázi obsahující záznamy vysílání Československé televize a webové rozhraní k ní. Databáze bude vycházet z naskenovaných stran *Týdeníku Československé televize*, které budou zpracovány OCR technologií. Naskenované strany byly pro účely této práce poskytnuty pracovníky archivu České televize. Primárním cílem práce je zpracování hlavního vysílání. Navíc bude v omezené míře zpracováno i lokální vysílání jednotlivých stanic pro Slovensko.

Naskenované strany z let 1966 až 1992 jsou ve vysokém rozlišení (cca 3400×5000 pixelů na naskenovanou stranu), což umožňuje snadné další zpracování. Před rokem 1966 nebyl vydáván Týdeník ČS televize, ale vysílání bylo součástí týdeníku s názvem *Československý rozhlas a televize*. Naskenované obrazy z něj bohužel nejsou dostatečně kvalitní (jak ilustruje příloha C) a pro tuto práci nemohly být použity.

Již v roce 2018 vytvořila Česká televize ku příležitosti 50. výročí událostí roku 1968 digitalizovanou verzi tehdejšího televizního vysílání [1], která je dostupná i online. Tento projekt slouží této práci jako vzor, z nějž bude vycházet webové uživatelské rozhrání.

V Týdeníku ČS televize se kromě televizního programu nacházejí i další části – různé rozhovory, inzerce a jiné. Tyto části nejsou v kontextu práce podstatné, předmětem zájmu jsou pouze naskenované soubory obsahující televizní program a zbytečné části je nutné odfiltrovat.

Jelikož by manuální přepisování programů byla práce na mnoho let, je vhodné celý proces patřičným způsobem automatizovat. Pro to bude využit systém OCR Tesseract, který je popsán v kapitole 2.2. Získané programy budou ukládány do relační databáze, jejímž návrhem, při němž bude zajištěna efektivita, se zabývá kapitola 3.1. Jak zmiňuje kapitola 2, OCR je proces sestávající z mnoha kroků, mezi kterými je přítomna silná diverzita. Kromě toho lze očekávat, že v rámci této práce bude zpracování obrazů systémem OCR pouze

jednorázová záležitost. Následné nahrání do databáze a tvorba webového rozhraní, přes které bude možné k databázi přistupovat, jsou další kroky, které jsou na sobě nezávislé.

Kvůli výše zmíněným důvodům nepovažuji za smysluplné vytvářet jeden velký program. Naopak mi ale dává smysl pro jednotlivé podproblémy, které je potřeba vyřešit, vytvořit samostatné nástroje. Výsledkem práce bude proto sada menších nástrojů, kdy výstup jednoho bude sloužit jako data pro zpracování nástrojem následujícím. Každý z těchto nástrojů bude spouštěn manuálně zvlášť. Společně pak budou nástroje zajišťovat splnění všech cílů potřebných pro dosažení výsledku. Dílčí kroky, kterých je nutno dosáhnout za účelem vypracování, je možné rozdělit následovně:

- Předzpracování obrazu pro zajištění vyšší kvality výsledků OCR
- Detekce určující, které z naskenovaných stran obsahují televizní program (za použití OCR)
- Extrakce pořadů, jejich uložení do dočasných souborů (opět s využitím OCR), dodatečná úpráva a nahrání do databáze
- Zpřístupnění databáze skrz jednoduché webové rozhraní

4.2 Analýza Týdeníku ČS televize

Během své existence prošel Týdeník ČS televize výraznými grafickými i obsahovými změnami. Na základě analýzy poskytnutých skenů bylo identifikováno celkem 8 základních variant televizního programu. Ty jsou k vidění v příloze B. Některé roky, ze kterých naskenované soubory pochází, byly skenovány ve formátu dvojstran (tzn. jeden soubor tvoří jak levá, tak pravá strana originálního Týdeníku). Ke zpracování odlišných variant bude nutno přistoupit jinak.

Kromě zmíněných variant struktury se televizní program lišil i obsahem. Původně obsahoval Týdeník ČS televize, kromě pro tuto práci irelevantních částí, pouze program vysílání pro jedinou televizní stanici, kterou Československá televize v té době nabízela. Přítomna byla i její lokální forma v podobě vysílání pro Slovensko – to však není primárním objektem zájmu této práce. Později bylo vysílání rozšířeno, konkrétně 10. května roku 1970 začala Československá televize vysílat na další stanici [17]. Původní stanice byla označena jako *1. program* a nová stanice jako *2. program*.

V tomto složení fungovalo vysílání dalších 20 let. Následně byla v květnu roku 1990 spuštěna stanice *OK3*, kde se vysílalo především zahraniční vysílání. K další změně koncepce došlo 3. září roku 1990 [11]. 1. program byl přejmenován na *F1*. V případě 2. programu došlo k rozdělení na českou a slovenskou část – *ČTV* a *S1* (později *STV*). V tomtéž roce se v televizním programu začaly objevovat i zahraniční stanice, které byly později, společně s programem stanice OK3, umístěny na vlastní stranu.

4.3 Editace obrazu

Před samotným zpracováním obrazu systémem OCR je vhodné obraz upravit do podoby, která zajistí přesnější výsledky. Procesem zvyšování kvality vstupního obrazu se zabývá kapitola 2.1.1. Prvním z implementovaných nástrojů bude takový, který zmíněné navýšení

kvality zajistí. Tento nástroj bude umožňovat prahování vstupních obrazů zvolenou metodou a dále provedení případných korekcí v závislosti na formátu televizního programu v daném roce.

Korekce mohou být například v podobě úpravy rotace nebo ořezání obrazu do podoby, která umožní systému OCR poskytovat lepší výsledky. V případě dvojstran budou s využitím tohoto nástroje strany rozděleny na dvě samostatné. Ořezávány budou i nedůležité části obrazů. Například v devadesátých letech měl každý den týdne v Týdeníku ČS televize svoji vlastní stranu. Televizní program nebyl jedinou věcí, co se na této straně nacházela. Kromě něj strana obsahovala i detailní popis vybraných pořadů, které však mohou narušovat práci systému OCR, a proto je vhodné tyto nepotřebné části odstranit. Dalším důvodem hrajícím ve prospěch odstranění je fakt, že ořezané strany, které budou oproti původním stranám menší a budou obsahovat méně informací, je systém OCR schopen zpracovat výrazně rychleji.

4.4 Detekce souborů s TV programem

V momentě, kdy jsou k dispozici vstupní obrazy ve formátu, který systém OCR dokáže zpracovat, může nastat samotný proces zpracování. Ještě před ním bude ale nutné vyhledat, které ze stran Týdeníku ČS televize program vysílání skutečně obsahují, neboť jak je popsáno v kapitole 4.2, vyskytují se v něm i jiné záležitosti. Získané soubory obsahující televizní program budou uloženy do textového souboru, což umožní případnou manuální korekci výsledků. Způsobů, jak program detekovat, je více.

Detekce klíčových slov identifikujících stranu s televizním program

Prvním z přístupů je detekování klíčových slov, která identifikují stranu s televizním programem. Již při skenování Týdeníku ČS televize pracovníky archivu České televize došlo k jeho rozdělení na jednotlivé roky. Tohoto faktu lze využít, jelikož v daném roce se obvykle formát programu neměnil a se znalostí roku je možné jednoznačně určit, jaká klíčová slova na jednotlivých stranách hledat. Tento přístup lze aplikovat nad obrázkem 4.1, který obsahuje horní část televizního programu z 1. ledna roku 1991.

Obrázek 4.1: Pravá horní část televizního programu z 1. ledna roku 1991 obsahující slova, která umožňují identifikovat stranu s televizním programem

Znázorněný formát byl (výjma názvu dnů, které se mění) totožný pro celý rok 1991. Na základě zpracování této části obrazu systémem OCR je možné identifikovat, zda soubor obsahuje televizní program. Toho lze docílit detekováním klíčových slov mezi výsledky zpracování obrazu systémem OCR. Jako klíčová slova mohou být v tomto případě použity názvy dnů, případně v kombinaci s datem, které je na obrázku taktéž možné vidět. Obdobný přístup lze použít i pro ostatní roky.

Kontrola přítomnosti pořadů

Druhý přístup, který lze zvolit, funguje na podobném principu jako přístup popsaný v předchozí části. Opět bude probíhat detekce klíčových slov, ovšem v tomto případě nebudou vyhledávána slova identifikující část s televizním vysíláním, ale jednotlivé televizní pořady.

Co mají pořady ve všech letech, kterými se tato práce zabývá, společné, je fakt, že jsou uvedeny vždy v totožném formátu, kdy první je uveden čas, jenž je následován názvem pořadu. Této skutečnosti využívá přístup, který je založen na kontrole, jestli výsledek zpracování malého výseku strany systémem OCR obsahuje určitý počet časů, které by indikovaly, že se jedná o televizní program. Na obrázku 4.2 je možné vidět malou část televizního programu z 1. ledna 1991, na základě které lze touto metodou soubor s televizním programem identifikovat.


17.20 — KRISTIÁN
Česká filmová veselohra o úředníkovi, který vede dvojí život. Scénář M. Frič, J. Gruss a E. Šimáček. Hrají: O. Nový, N. Gollová, B. Veverka, R. Schránil, J. Marvan, Č. Šlégl a další. Režie M. Frič
18.55 — Reklama
19.00 — Večerníček
Jája a Pája
19.10 — Přehled večerních pořadů
19.20 — Předpověď počasí
19.25 — Reklama
19.30 — Deník ČST
20.10 — Nešťastná šťastná Hana Hegerová
Hudební medailón. Připravili F. Fenič a V. Polesný
(Premiéra — Bratislava)

Obrázek 4.2: Část televizního programu pro 1. ledna 1991 s časy vysílání, která umožňují detekovat soubor s televizním programem

Klady a zápory jednotlivých přístupů

Oba zmíněné přístupy mají své klady i zápory. Mezi zápory patří v případě prvního přístupu to, že může dojít k nesprávnému načtení klíčového slova systémem OCR, což bude mít za následek, že strana nebude jako program identifikována.

V případě druhého přístupu lze zase narazit na problém, že televizní pořady nebyly přítomny pouze v části s televizním programem pro Československou televizi, ale například i v části obsahující vysílání zahraničních satelitních stanic, na které tato práce zaměřena není. To znamená, že druhý přístup vytváří opačný problém než přístup první – jako strany s televizním programem mohou být detekovány i ty, které jej ve skutečnosti neobsahují.

4.5 Extrakce pořadů

Po detekci, které z obrazových souborů obsahují televizní program, bude dalším krokem tento program extrahovat. K tomu bude použit, stejně jako u detekce programů, systém OCR. Již během extrakce pořadů bude nutné vhodným způsobem určit, jaké měl pořad údaje – název, případný popis a datum a čas vysílání.

Výsledkem fáze extrahování pořadů bude jejich uložení do textových souborů. To považuji za vhodné, aby mohlo v případě potřeby dojít pouze k úpravám již získaného textu a časově náročný proces zpracování systémem OCR nemusel být opakován.

Formát televizního programu

Jak je zmíněno v kapitole 4.2, televizní program nabýval v průběhu let mnoha grafických podob. Formáty TV programu se lišily například v počtu dnů vysílání na stránce. V různých letech bylo na jednu stranu umísťováno odlišné množství programů. Zatímco v prvních letech, kterými se tato práce zabývá, se program na celý týden vešel na jednu stranu, s rozšiřováním televizního vysílání začal televizní program zabírat místa více. Následovalo rozšíření týdenního programu na dvě strany a již v osmdesátých letech byla pro každý den vyhrazena vlastní strana. K poslední změně došlo v letech devadesátých, kdy byly pro jeden den vyhrazeny strany dvě.

V rámci grafického uspořádání se též měnilo umístění televizního programu lokálního slovenského vysílání, které bylo v určitých letech dokonce z programu Československé televize odsunuto na jinou stranu. Vysílání pro Slovensko však není hlavním předmětem zájmu této práce, a proto v letech, kdy nebylo součástí standardního televizního programu, zpracováno nebude. Další věcí, která se v televizním programu měnila, byla pozice vysílání dalších stanic. Se všemi těmito fakty bude nutno při implementaci programu počítat.

Získané pořady budou rozděleny na tři části – čas vysílání, jejich název a případný popis. Názvy pořadů byly vždy psány tučným písmem. Nabízela by se možnost tučné písmo detekovat, což ovšem systém OCR Tesseract nepodporuje. Jelikož byly názvy pořadů často dostatečně krátké a vešly se na jeden řádek, je možné jako název pořadu určit vždy celý řádek, který následoval za časem vysílání. Zbytek textu bude tvořit popis pořadu.

Dodatečná úprava výsledků

Výsledky, které systém OCR vyprodukuje, nemusí být perfektní, proto bude vhodné je dále upravit. Některými postupy, jak toho docílit, se zabývá kapitola 2.1.5. Součástí dodatečné úpravy bude validace, že získané pořady jsou správné. Jednou ze situací, co může nastat, je to, že systém OCR není schopen z mnoha možných důvodů rozpoznat pořad a detekuje pouze jeho část. V takovémto případě budou nevalidní pořady odstraněny. Další z problémů, co může nastat, je spíše kosmetičtějšího rázu. Může se stát, že jedna mezera bude systémem OCR zpracována jako mezery dvě či více. Dalším problémem může být přiřazování textu k pořadům, ke kterým nepatří. V získaných slovech mohou být také chyby, což je vhodné opravit automatickou korekcí, např. podle slovníku. Upravené pořady budou následně opět uloženy do textového souboru.

Nahrání programu do databáze

Poslední částí procesu extrakce pořadů bude jejich nahrání do databáze. K tomu budou využity textové soubory, které jsou výsledkem předchozí etapy.

Během tohoto procesu bude nutné načtené řádky vhodně rozdělit. Jednotlivé řádky se skládají z více částí – konkrétně se jedná o čas vysílání pořadu, jeho název a popis, jehož součástí mohou být osoby, které se na pořadu podílely. Ty jsou uvedeny včetně rolí, které při přípravě pořadu sehrály. V rámci této fáze bude nutné tyto osoby v popisu pořadu vyhledat. Pro to lze využít opět způsob detekování klíčových slov, kterými zde budou názvy rolí. Tento případ ilustruje obrázek 4.3. Na obrázku je možné vidět osobu odpovědnou za scénář, režiséra a poté jednotlivé herce v hlavních rolích. Po detekování jednoho z klíčových slov v podobě role bude nutné pokusit se detekovat jméno osoby. V případě, že bude jméno úspěšně detekováno, bude vloženo do databáze včetně role, kterou osoba v daném pořadu zastávala.

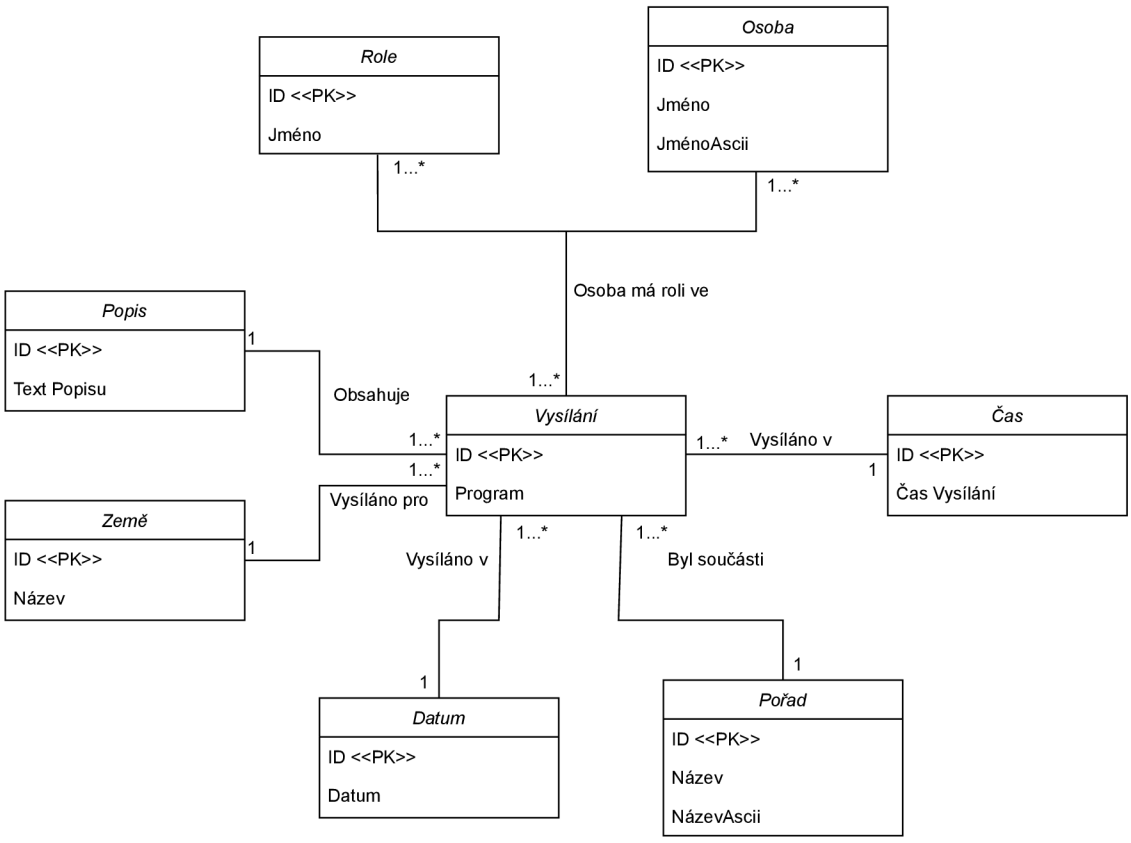

Francouzský televizní film na motivy románu E. Zoly. Scénář C. Brule. Režie J. Rouffio. Hrají: Miou—Miou, C. Brasseur, M. Galabru, A. Galienová, J. P. Bisson, R. Rimbaud a další. Režie slovenského znění E. Balgová— Harantová

Obrázek 4.3: Ukázka záznamů osob, co se na pořadu podílely, a jejich role v pořadu – to vše se nástroj pokusí do databáze uložit

4.6 Návrh relační databáze

Nejvýznamnější částí této práce je, jak již samotný název práce napovídá, databáze vysílání pořadů. V této databázi budou ukládána všechna získaná data. Jejím efektivním návrhem se zabývá kapitola 3.1. Lze očekávat, že množství pořadů, které budou systémem OCR načteny, bude velké. Z toho důvodu by bylo vhodné, aby databáze neobsahovala zbytečné informace, které by zvětšovaly její velikost.

Na základě podrobné analýzy televizního programu a požadavků na tuto práci byla navržena relační databáze, která je schopna ukládat všechny potřebné informace. Popis databáze ilustruje ER (entity relationship) diagram, jenž je možné vidět na obrázku 4.4.



Obrázek 4.4: ER diagram reprezentující návrh databáze

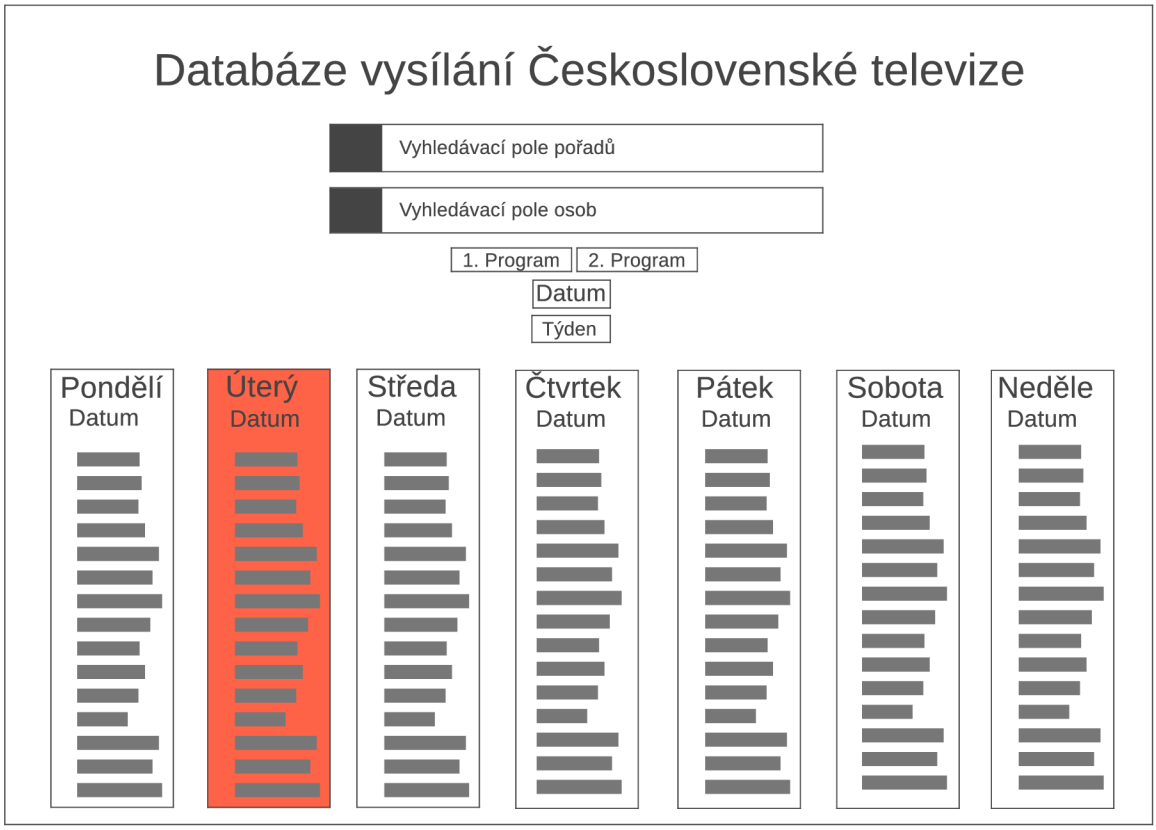
Databáze umožní mezi uloženými pořady vyhledávat a také je filtrovat podle data. Kromě toho nabídne též možnost vyhledávat i jednotlivé osoby, které se podílely na přípravě různých televizních pořadů. Za tímto účelem bude databáze obsahovat kromě jmen osob či

názvů pořadů i jejich verze v ASCII formátu, aby bylo možné mezi nimi vyhledávat i bez použití diakritiky a vyhledávání tak bylo intuitivnější.

4.7 Webové rozhraní

Poslední částí procesu tvorby databáze vysílání ČS televize je vytvoření jednoduchého webového uživatelského rozhraní. Uživatelské rozhraní bude postaveno na základě projektu České televize dostupném z [1].

Oproti vzoru, ze kterého rozhraní vychází, v něm bude přítomno více elementů. Ty slouží převážně pro vyhledávání v databázi. Konkrétně se jedná o vyhledávací pole, která slouží pro vyhledávání osob a pořadů. Tato pole obsahují funkci automatického našeptávání názvů pořadů na základě vloženého textu, s čímž počítá i návrh relační databáze, jak je popsáno v předchozí sekci 4.6. Další věcí, kterou bude rozhraní obsahovat, bude kalendář, který bude sloužit pro výběr data vysílání, které si chce uživatel zobrazit. Ilustraci návrhu rozhraní v podobě drátěného modelu je možné vidět na obrázku 4.5. Obrázek ilustruje i zvýraznění vybraného dne. Ten koresponduje se dnem, který byl zvolen v kalendáři či případně zapsán manuálně.



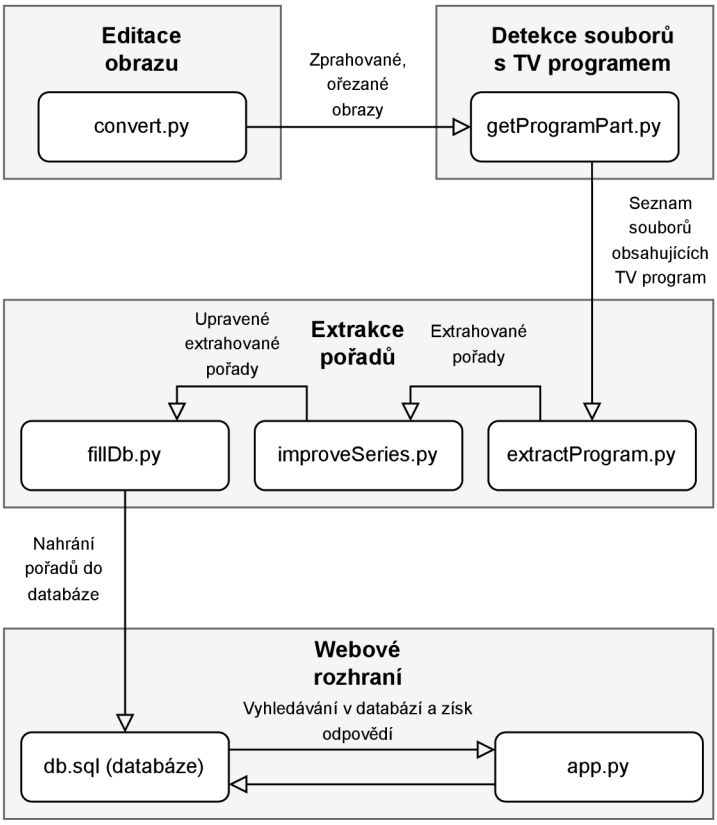
Obrázek 4.5: Drátěný model reprezentující návrh jednoduchého webového uživatelského rozhraní se zvýrazněním vybraného data

Kapitola 5

Implementace

Tato kapitola se zabývá implementací sady nástrojů, jejichž základní návrh byl popsán v předchozí kapitole 4. Popsány zde budou stěžejní části implementace jednotlivých nástrojů a také bude představen formát souborů, do kterých jsou ukládány mezivýsledky dílčích částí řešení.

Implementované nástroje byly pro přehlednost a oddělení konkrétních činností oproti návrhu dále rozděleny. Na obrázku 5.1 je možné vidět konkrétní soubory a jejich vzájemné interakce. Jednotlivé nástroje jsou spouštěny manuálně po sobě v sekvenci, jakou obrázek ilustruje.



Obrázek 5.1: Schéma soustavy nástrojů

Webové rozhraní obsahuje kromě hlavní řídící aplikace `app.py` mnoho dalších pomocných souborů jako jsou například HTML soubory a nebo další pomocné skripty.

5.1 Použité technologie

Před samotným popisem implementace jednotlivých součástí řešení budou představeny technologie, které jsou pro jejich implementaci použity. Jako programovací jazyk byl zvolen **Python3**[1], konkrétně ve verzi 10.4. Implementace probíhala na operačním systému Windows 10 Pro. Další použité technologie jsou popsány v následujících podkapitolách.

Knihovna OpenCV

Důležitou součástí implementace je knihovna OpenCV[2], konkrétně pak její verze pro jazyk Python – **opencv-python**[3]. Jedná se o knihovnu sloužící primárně pro účely počítačového vidění a strojového učení. Je implementována v jazyce C++. Pro účely této práce jsou využívány pouze její části sloužící pro práci s obrazovými soubory. Konkrétní využití nachází knihovna ve všech částech této práce, které pracují se sytémem OCR, kdy je používána pro načtení obrazů, které jsou systému OCR poskytnuty. Další využití má při předzpracování obrazu, kdy jsou využity její algoritmy pro prahování a ořezávání.

Tesseract

Snad nejdůležitější součástí celé práce je systém OCR Tesseract. Jeho popis se nachází v sekci 2.2. Jedná se o systém s pravděpodobně největší uživatelskou základnou ze všech volně dostupných systémů OCR. Jeho výsledky jsou poměrně kvalitní a pro účel extrakce pořadů dostačující. Nebyl používán přímo originální Tesseract, neboť ten je vyvíjen v jazyce C/C++, ale tato práce využívá jazyk Python. To bylo vyřešeno využitím Python obalu nad tímto nástrojem. Konkrétně je použit modul **pytesseract**[4].

SQLite

Jak bylo předesláno již při představování existujících databázových systémů, jako databázový systém slouží SQLite, který je popsán v kapitole 3.2. Do vytvořené databáze jsou ukládány načtené pořady a další s nimi související záležitosti.

Knihovna SQLAlchemy

Pro práci s databází, konkrétně pro nahrávání jednotlivých záznamů a posléze pro vyhledávání v ní, slouží knihovna SQLAlchemy[5]. Jedná se o knihovnu napsanou pro jazyk Python za účelem práce s mnoha různými databázovými systémy. Umožňuje uživatelům velmi jednoduše přistupovat k datům v databázi a nabízí plnou podporu jazyka SQL, přístupnou skrz jazyk Python.

[1] https://www.python.org/
[2] https://opencv.org/
[3] https://pypi.org/project/opencv-python/
[4] https://pypi.org/project/pytesseract/
[5] https://www.sqlalchemy.org/

Flask

Pro vývoj jednoduchého webového uživatelského rozhraní byl zvolen Flask[6]. Jedná se o jednoduchý webový aplikační rámec, který slouží primárně pro rychlé prototypování za účelem efektivního vývoje aplikace, kterou je možné jednoduše upravovat. V základu se jedná o *mikro* rámec, který nepodporuje abstrakci nad databází, validaci formulářů a jiné. Přestože tyto součásti nejsou zahrnuty v jádru rámce, je možné je přidat jako dodatečné moduly. Uživatel tak není omezen ve výběru, jaké databázové rozhraní či jiný modul si zvolí. Jelikož se modul SQLAlchemy používá již při načítání dat do databáze, je použit i společně s Flaskem pro operace nad databází zahrnující v kontextu uživatelského rozhraní pouze její čtení. Pro vykreslování stránek jsou ve Flasku použity tzv. šablony (*templates*), které do připravených HTML souborů přiřazují konkrétní data.

5.2 Nástroj pro přípravu obrazu

Jak bylo již zmíněno, Týdeník ČS televize byl už při skenování rozdělen na jednotlivé roky, ze kterých pochází. Formát Týdeníku se však neměnil pouze po skončení roku, ale občas i v průběhu. Proto bylo nutné naskenované strany dále rozdělit na jednotlivé formáty, což bylo provedeno manuálně. Nástroj, který se předzpracováním obrazu zabývá, je `convert.py`. Umožňuje spustit vybraný prahovací či ořezávací algoritmus buď na konkrétní soubor, nebo také na všechny soubory ve vybrané složce.

Prahování

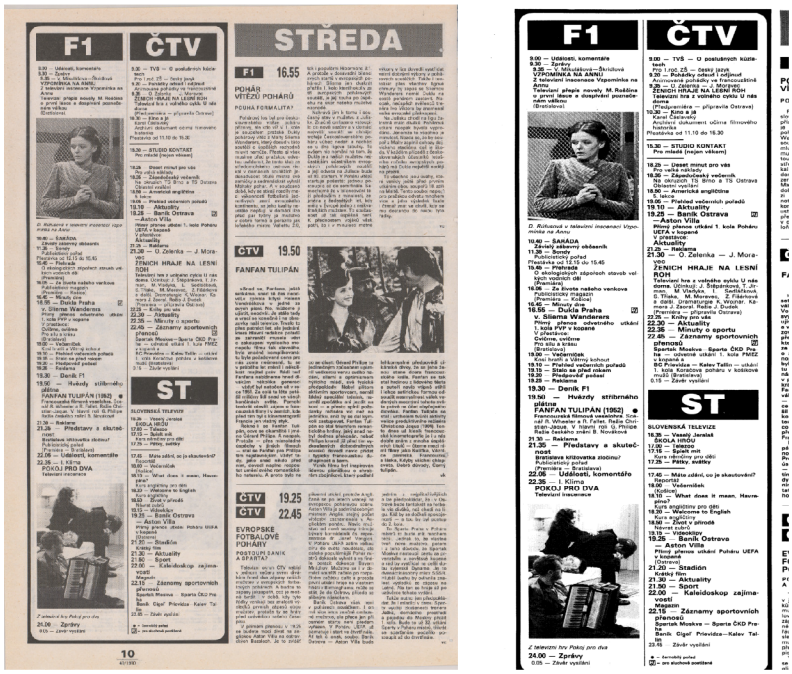
Bylo již řečeno, že pro operace s obrazy slouží knihovna OpenCV. Své využití nachází hned v prvním kroku při prahování obrazu. Při implementaci prahovacích algoritmů jsem vycházel z návodu[7], který pochází ze stránek OpenCV. Pro přípravu většiny obrazů se ukázalo být nejvhodnější globální prahování. Adaptivní prahování nevytvářelo v některých případech uspokojivé výsledky (což vedlo k nefunkční automatické segmentaci textu) a Otsuho metoda často z důvodu jenom mírně odlišného stínu v různých částech naskenovaného papíru zvolila práh, který vedl k horšímu výsledku, než jakého bylo možné dosáhnout použitím pevného globálního prahu. Ten byl vždy volen na základě testování na vybraných snímcích pro daný formát, kdy nejkvalitnější výsledky určily, jaký práh bude použit.

Ořezávání obrazu

Další funkcí, kterou nástroj nabízí, je ořezávání obrazů. Jak je popsáno již v kapitole 4.3, na straně, která obsahuje program vysílání pro daný den, se občas nachází zbytečné informace. V případě, kdy jsou tyto informace ponechány, dochází k problémům souvisejícím s nesprávně provedenou automatickou segmentací textu a zpracování strany trvá déle. Pro vyřešení těchto problémů a zajištění vyšší kvality výsledků jsou nadbytečné informace odstraňovány. Jelikož měl televizní program vždy v konkrétním časovém období totožný formát, je možné pevně zvolit bod, ve kterém má být obraz ořezán. Ten je volen manuálně. Totožně je řešen i problém s dvojstranami v letech, kdy dvojstrany obsahovaly na každé své straně program pro odlišný den – takovéto soubory jsou rozděleny na dva a každý z nich obsahuje pouze program pro konkrétní den. Proces ořezávání byl vždy spouštěn stejným

[6] https://flask.palletsprojects.com
[7] https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html

způsobem na všechny soubory z daného roku, respektive období, kdy byl televizní program ve stejném formátu. Ilustraci účelu tohoto nástroje je možné vidět na obrázku 5.2. Nalevo se nachází originální vstupní obraz a napravo už zprahovaný a ořezaný obraz.



(a) Původní obraz

(b) Ořezaná část obsahující pouze televizní program

Obrázek 5.2: Srovnání vstupních obrazů před a po prahování a ořezání nepodstatných informací

Přestože ořezaný obraz 5.2 stále obsahuje zbytek původního textu, již to nepředstavuje výrazný problém, neboť je možné jej odstranit v pozdější fázi při dodatečném zpracování výsledků. Samotná přítomnost zbytkového textu je zapříčiněna tím, že skenování původního programu neprobíhalo uniformně a jednotlivé naskenované obrazy mohly být vždy o kousek posunuty. Při ořezávání bylo toto nutné vzít v potaz a raději zvolit bod ořezání tak, aby se občas objevovaly zbytkové části původního obrazu, než aby byla nesprávně ořezána i část samotného televizního programu.

Rozdíl oproti návrhu

Původní návrh počítal i s implementací funkce na korigování nesprávné rotace. V souvislosti s tím se však vyskytl problém, kdy některé z obrazů tvořících dvojstrany byly natočeny odlišně a korekce k ničemu nevedla. Další problém, který s jejím využitím nastal, vycházel z přítomnosti velkého množství různých obrázků v televizním programu. Na těch byly nejčastěji ukázky z pořadů či fotky hlasatelů. Systém OCR Tesseract je ale schopný pracovat i s mírně narotovaným obrazem a stále poskytnout správné výsledky, což znamená, že bez této funkcionality je možné se obejít.

5.3 Nástroj pro detekování souborů s TV programem

V předzpracovaných souborech je nutné rozlišit, které z nich obsahují televizní program. Vyhledávání souborů s TV programem řeší nástroj `getProgramPart.py`. V kapitole 4.4, věnující se návrhu detekování, jsou popsány dva způsoby, jak k tomuto procesu přistoupit. Ve správných podmínkách dosahují dobrých výsledků oba dva. V případě, kdy nejsou klíčová slova na straně jasně rozlišitelná systémem OCR, neboť by bylo zapotřebí zvolit jiné metody prahování z důvodu barevné odlišnosti klíčových slov v průběhu roku, první metoda selhává. Pro finální implementaci byla proto zvolena metoda druhá, tzn. dochází k detekci přítomnosti určitého množství časů vysílání.

Ořez strany

Jádrem programu je řada funkcí, které slouží pro dočasné ořezání jednotlivých vstupních obrazů. Spouštět systém OCR na celý obraz, v němž je přítomno mnoho textu, by bylo pomalé a hlavně zbytečné – v takové situaci už by mohlo rovnou dojít i k extrakci pořadů a doba běhu programu by se zněkolikanásobila. To celé by bylo ještě horší v případě, kdy by jako program byly označeny i nesprávné soubory. Právě tento problém řeší ořezávání obrazů. Podstatné je počítat s případy, kdy může technika detekce selhat. To se může stát mimo jiné kvůli posunům mezi jednotlivými naskenovanými obrazy, o kterých padla zmínka už v předchozí kapitole v souvislosti s výběrem vhodného bodu řezu. Tyto případy jsou řešeny redundancemi za podmínky, že byla první detekce neúspěšná. V takovém případě dojde k ořezání obrazu na jiné pozici a celý proces se zopakuje. Jestliže ani v tomto případě nebyl televizní program nalezen, značí to, že se v daném souboru pravděpodobně nenachází.

Detekce časů vysílání

Poté, co je k dispozici ořezaná část obrazu, je nutné zjistit, zda její obsah indikuje přítomnost televizního programu. Nejvhodnějším způsobem, jak toho docílit, se ukázala být kontrola přítomnosti časů vysílání popsaná v kapitole 4.4. Čas samotný totiž znamená, že se za ním s velkou pravděpodobností nachází název pořadu. V získané části strany jsou systémem OCR detekována slova. Časy jsou taktéž rozpoznány jako samostatná slova. Celkový výsledek analýzy, který je v podobě pole s textem, stačí otestovat na přítomnost konkrétního množství časů. Pro tento přístup je používán regulární výraz, vůči kterému je každé slovo v seznamu získaných slov porovnáno. Tvar regulárního výrazu byl určen na základě analýzy hodnot výsledků, které systém OCR Tesseract z detekovaných časů nejčastěji vracel, a regulární výraz proto počítá i s nejčastějšími chybami, ke kterým mohlo dojít. Kontrola času je zajištěna funkcí, která je k vidění jako výpis 5.1.

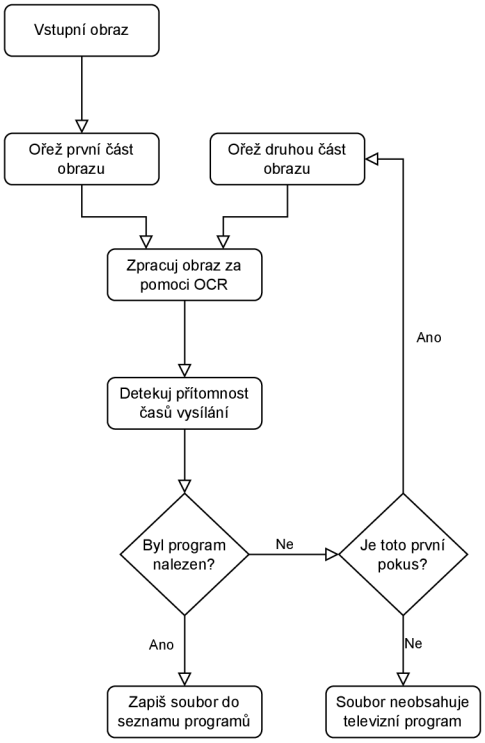
```
1 def isTime(word):
2     regex_string = "^( |-|„|\()?[0-9]+(\.|,) ?[0-9]+:?\
3 ((-|-)[0-9]+(\.|,) ?[0-9]+:?)?\.?$"
4     return re.search(regex_string, word):
```

Výpis 5.1: Funkce pro detekci, zda je slovo časovou hodnotou

Po zpracování celého textu je zkontrolováno, zda byly nalezeny alespoň *3* časy vysílání. V takovém případě je název souboru, včetně celé cesty k němu, uložen do textového souboru, který obsahuje jména všech souborů obsahujících televizní program pro daný rok, případně pro část roku obsahující totožný formát TV programu. To umožňuje manuální kontrolu nad výsledky a provedení případného odstranění falešných pozitiv či přidání souborů, které

se naopak lokalizovat nepodařilo. Před dalším krokem je nutné, aby všechny ze získaných souborů skutečně obsahovaly televizní program, protože kdyby toto splněno nebylo, v dalším kroku by to vedlo k problémům.

Celý postup detekce, zda soubor obsahuje program vysílání, ilustruje vývojový diagram, který se nachází na obrázku 5.3.



Obrázek 5.3: Vývojový diagram reprezentující proces, jakým probíhá detekce, zda je součástí souboru televizní program

5.4 Nástroj pro extrakci pořadů

Jednotlivé pořady jsou detekovány pouze z obrazů, které byly označeny jako obsahující televizní program. Extrakce pořadů je řešena nástrojem `extractProgram.py`. Na daných souborech je spuštěn systém OCR Tesseract, stejně jako v předchozím kroku. V souvislosti s ním je vhodné názorně ukázat, jak toto OCR zpracování vypadá. To je možné vidět na výpisu 5.2.

```
1    TESSERACT_CONFIG = r'-l ces --psm 1 --oem 1'
2    img = imread(file)
3    ocr_results = pytesseract.image_to_data(img,
4                  output_type=Output.DICT, config=TESSERACT_CONFIG)
5    text = ocr_results.get('text')
```

Výpis 5.2: Spouštění Tesseractu s automatickou segmentací

Ve výše znázorněném případě je Tesseract spouštěn nad obrazovým souborem načteným funkcí `imread` pocházející z knihovny OpenCV. Je použita metoda `image_to_data`, která slouží k získání všech součástí načteného textu – textu samotného, segmentovaných bloků, míry přesnosti, s jakou byly jednotlivé znaky segmentovány, a dalších. Výstup je

nastaven parametrem `output_type`, zde jako slovník. Důležitou součástí volání funkce je nastavení Tesseractu, které má být v tomto volání použito. Řetězec `TESSERACT_CONFIG` definuje formát parametrů, které jsou systému OCR předávány.

V případě, kdy je Tesseract nainstalován a správně umístěn v systémové proměnné PATH, je možné si všechny možnosti parametrů zobrazit v příkazové řádce příkazem `tesseract --help-extra`. `-l ces` je parametr sloužící pro specifikaci jazyka, ve kterém je text na zdrojovém obrazu napsaný. Tento parametr zde nabývá hodnoty `ces`, která reprezentuje češtinu. `--psm 1` je způsob, jakým má být přistupováno k segmentaci vstupního obrazu na jednotlivé bloky textu. Hodnota 1 značí, že segmentace má být automatická a zůstává čistě v roli Tesseractu. `--oem 1` pak značí, že má být použito jádro Tesseractu využívající neuronové sítě.

Kód na výpisu 5.2 vrací získaný text, který byl již Tesseractem segmentován do bloků. To znamená, že v ideálním případě je v této chvíli k dispozici pole obsahující slova, která po sobě následují tak, jak šla za sebou v originálním textu – řádky textu byly odděleny a teprve po přečtění prvního sloupce došlo ke zpracování druhého.

Určení data vysílání

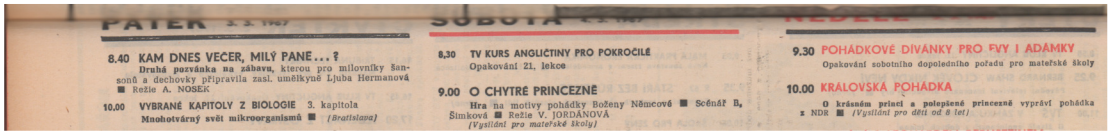
U získaného televizního programu je nutné určit datum, ke kterému patří. To není vždy jednoduše možné. Byly případy, kdy bylo datum uvedeno výrazně jasnější barvou než okolní text a bylo v procesu prahování ztraceno. Tento případ ilustruje obrázek 5.4. Problém nebylo možné spolehlivě řešit ani využitím adaptivního prahování, neboť jeho výsledky nedokázal systém OCR zpracovat a nebyl ve výsledku oproti globálnímu prahování rozdíl.

(a) Původní televizní program obsahující označení dne vysílání

(b) Chybějící den vysílání po provedení procesu prahování

Obrázek 5.4: Ukázka ztráty dne vysílání po provedení procesu prahování

Dalším problémem, který byl způsoben grafickým designem Týdeníku ČS televize, bylo, že datum nebylo na daném programu vůbec a bylo umístěno až na následující straně. Tak tomu bylo například v říjnu roku 1990. Posledním případem, který se objevoval už na originálních naskenovaných obrazech, bylo, že datum nebylo čitelné z důvodu nekvalitních zdrojových materiálů či nevhodného procesu skenování. Tento případ ilustruje obrázek 5.5.



Obrázek 5.5: Ukázka nepřítomnosti data z důvodu nevhodně naskenovaného obrazu

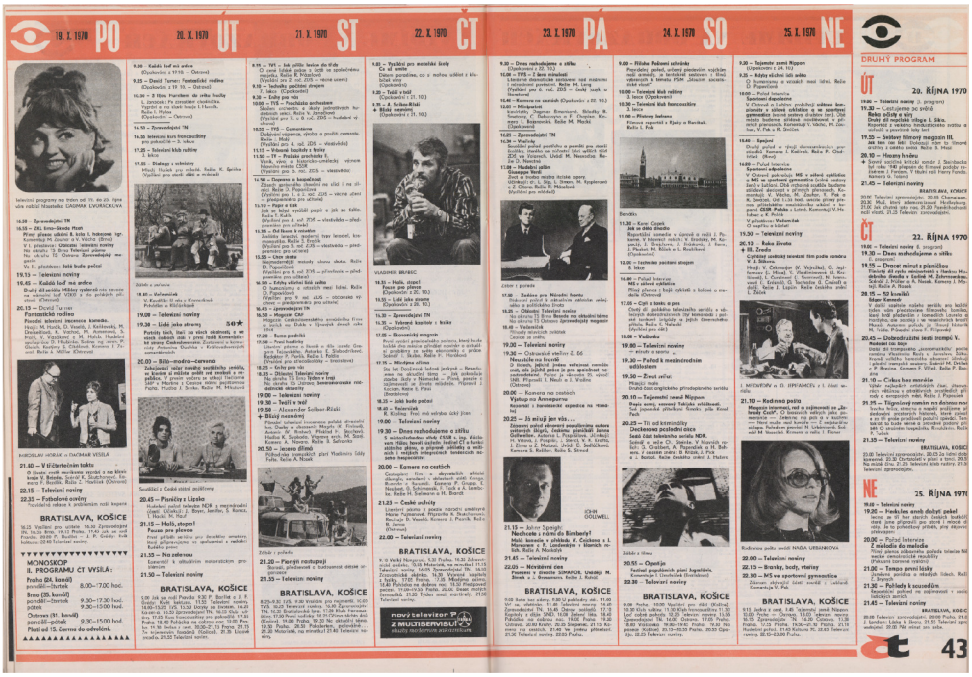
Spoléhat na detekci konkrétního klíčového slova pro určení data vysílání by ani v nejlepším případě, kdy byla data přítomna i po procesu prahování, nebylo ideální z důvodu, že toto datum nemusí být systémem OCR detekováno správně.

Řešením, které se ukázalo dostačujícím, je zadávat datum manuálně. V ideálním případě byl televizní program pro jeden den umístěn pouze v jednom souboru. Tehdy stačí po

zpracování souboru datum pouze inkrementovat. Takto ideální situace však nastává pouze v pozdějších letech. V dřívějších letech byl na jedné straně program na celý týden. V takovémto případě je úvodní datum zvoleno opět manuálně, a vždy po detekci konce dne dojde k jeho inkrementaci. Způsob této detekce je detailněji vysvětlen níže.

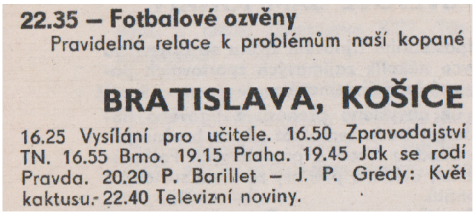
Datum vysílání v případě více programů v souboru

Vysvětlena bude situace na programu z roku 1970, konkrétně z 19.-25. října, který je možné vidět na obrázku 5.6.



Obrázek 5.6: Ukázka programu z roku 1970

Daný program je rozdělen do sloupců, což dokáže systém OCR detekovat a správně sloupce segmentovat. Na konci sloupce dojde k detekci času vysílání posledního pořadu pro daný den – tuto část vysílání ilustruje obrázek 5.7, který zobrazuje konec pondělního vysílání z obrázku 5.6.



Obrázek 5.7: Televizní program ilustrující konec vysílání ze dne 19. 10. 1970

Časem vysílání posledního pořadu na obrázku 5.7 je `22.35`. Hodina vysílání je uložena a při detekci dalšího času jsou hodiny porovnány. Zde je hodina nižší (`16.25`), což značí, že daný program už je součástí nového dne, nebo patří pod jinou stanici. V takovém případě, jestliže se nejedná o Slovenské vysílání nebo 2. program, je navýšeno počítadlo dnů.

Ve zmíněné situaci se ale jedná právě o lokální vysílání pro Slovensko. Jak je možné vidět na obrázku 5.7, to je označeno klíčovými slovy *BRATISLAVA, KOŠICE* a nachází se na konci sloupce. Tato slova je nutné detekovat. V případě, že dojde ke snížení času vysílání, musí nejprve dojít ke kontrole, zda tento program není součástí vysílání pro Slovensko pro daný den. V případě, že ano, navýšení dne neproběhne. Pro zvýšení přesnosti jsou před kontrolou písmena vždy převedena na velká. Obdobný přístup je použit i u části programu obsahující monoskop, na obrázku 5.6 k vidění vlevo dole.

Pokud nebyl detekován ani monoskop, ani vysílání pro Slovensko, dojde, jak už bylo řečeno, k navýšení čítače dnů a další program je již označen jako patřící do dalšího dne. Stejně byly zpracovány i ostatní sloupce. Televizní program na obrázku 5.6 ovšem obsahuje i program vysílání 2. programu. V případě, že je detekováno klíčové slovo *DRUHÝ PROGRAM*, je v tomto konkrétním případě program označen jako součást vysílání 2. programu a datum je nastaveno na hodnotu přítomnou hned pod ním.

Podobný přístup je použit i v případě, kdy je na jednom vstupním obrazu pouze program pro jeden den. V takovém případě snížení vysílacího času pořadu neznamená, že by byl pořad součástí dalšího dne, ale že je součástí vysílání další stanice a podle toho je označen. V takovém případě jsou dny, jak již bylo zmíněno, inkrementovány vždy po zpracování jednoho souboru.

Určení názvu pořadu

Způsob detekce názvu, který se v praxi ukázal nejvhodnější, využívá vlastností Tesseractu pro určení textu patřícího na jeden řádek. Tesseract po načtení řádku přidá několik prázdných znaků reprezentujících prázdné místo, než opět detekuje další řádek. Pro určení názvu pořadu je využíváno těchto prázdných míst. V případě, že byl již načten čas, je detekováno, zda nebylo nalezeno prázdné místo. Jestliže ano, dojde k přidání řetězce *NAZEVPORADU* jako následujícího slova do načteného pořadu. To slouží k oddělení názvu a popisu. Výsledek detekce názvu může vypadat například takto:

```
8.50 Zprávy NAZEVPORADU
```

Obecně lze tento přístup shrnout tak, že jako název pořadu slouží všechen text, který následuje po čase vysílání na stejném řádku. Jelikož nelze název jednoduše určit algoritmicky, tento přístup se ukázal být dostačující alternativou. Nevýhodou je, že v případě, kdy se na stejném řádku nacházel i popis, tento způsob přidá i zmíněný popis jako součást názvu. Podobný problém nastává i tehdy, když byl název rozdělen na více řádků – v tom případě je součást názvu z dalších řádků určena jako součást popisu. Výsledky byly v těchto případech sice nesprávné, ale celkově akceptovatelné.

Ukládání načtených programů

Všechny detekované informace jsou po zpracování každého souboru přidány do textového souboru obsahujícího program pro daný rok či část roku. To je prováděno hlavně za účelem možnosti manuální korekce výsledků, která je v případě neúspěchu automatické segmentace v podstatě nutná. Dalším důvodem je, že tento přístup umožňuje opakované nahrávání do databáze, aniž by muselo dojít k opětovnému zpracování obrazů.

Formát, ve kterém jsou extrahované pořady ukládány, je následující:

```
[Země vysílání]
[Datum vysílání ve formátu D. M. YYYY]
[Stanice, na které se pořady vysílaly]
[Čas vysílání] [Název pořadu] [Popis pořadu]
```

Hlavička je zapsána vždy po detekci nového dne či stanice. Záznam pořadů mohl vypadat například tak, jak je ukázáno níže:

```
CZ
31. 12. 1990
1
8.30 SILVESTR PRO DĚTI NAZEVPORADU Pořad Studia Kamarád
10.00 Ze světa kouzel NAZEVPORADU Setkání s~mistry magie
```

5.5 Nástroj pro automatickou úpravu výsledků

K navýšení kvality výsledků slouží nástroj `improveSeries.py`. Vstupem nástroje jsou textové soubory obsahující extrahované pořady. Výstupem jsou opět textové soubory s extrahovanými pořady, tentokrát však v upravené podobě.

Primárním účelem tohoto nástroje je odstranit nevalidní pořady či popisy a upravit další záležitosti. Prvním krokem, který tento nástroj vykonává, je kontrola, zda je načtený řádek s pořadem validní. V případě, že je součástí řádku pouze část hlavičky obsahující metadata jako datum a vysílací stanici, je řádek z hlediska zpracování ignorován a dojde pouze k jeho přidání do výsledného souboru. V případě, že řádek metadata neobsahuje, dojde jako první k jeho validaci.

Validace pořadu

Validace spočívá v kontrole, jestli byl pořad správně načten – je kontrolována přítomnost času vysílání a názvu pořadu. Název pořadu musí pro svou validitu splňovat pouze jedno kritérium, a to obsahovat alespoň tři nebílé znaky. Této hodnoty bylo dosaženo na základě analýzy pořadů, které se během let vysílaly, a faktu, že při této analýze nebyl nalezen žádný pořad, co by kritérium nesplňoval. V případě chyby v systému OCR, kdy došlo k nesprávné segmentaci či rozpoznání písmen, tento problém ale mohl nastat. Takový pořad je vyhodnocen jako neplatný a není zařazen do výsledků.

Úprava času vysílání

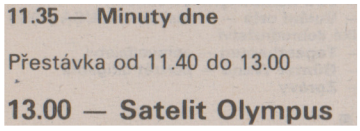
Po kontrole validity názvu pořadu dojde k jeho editaci. Prvním krokem, který je v rámci této editace proveden, je unifikace formátu času vysílání. Ten se v televizním programu nacházel ve formátu `H.MM`, například `9.30`. Pro účely dalšího zpracování je tento čas převeden do formátu `HH.MM`, v tomto případě `09.30`.

Úprava popisu pořadu

Dalším krokem je úprava popisu pořadu. Primárním cílem je odstranit nevalidní popisy. Ty se objevovaly v případě, kdy Tesseract nesprávně segmentoval stranu a přiřadil do bloku i text, který k němu opticky nepatří. Tento text je ve výsledcích OCR zpracování zpravidla oddělen od korektního textu mezerami reprezentujícími prázdné místo, které Tesseract

interpretoval taktéž jako součást bloku textu. Tento problém je řešen kontrolou, zda popis pořadu neobsahuje více než tři mezery následované alespoň jedním dalším znakem. V případě, že ano, je celý popis ořezán a zachována je pouze část nacházející se před mezerami. Tři mezery byly zvoleny opět na základě analýzy výsledků poskytnutých Tesseractem, kdy právě tři a více mezer následované dalším znakem indikují, že daný text k popisu pořadu nepatří.

S úpravou pořadu dále souvisí odstranění textu indikujícího přestávku ve vysílání. Jestliže je v popisu pořadu detekováno klíčové slovo *Přestávka* (nebo jeho varianty plynoucí z případného nesprávného rozpoznání), je opět celá část za tímto slovem, včetně slova samotného, odstraněna, neboť značí přestávku ve vysílání a nikoliv popis pořadu. Ukázka přestávky ve vysílání, která nepatří do popisu pořadu, je k vidění na obrázku 5.8.



Obrázek 5.8: Přestávka ve vysílání, která nepatří do popisu pořadu a je proto z textu odstraňována

Posledním krokem je odstranění přebytečných mezer. To se týka jak názvu pořadu, tak i jeho popisu. V případě, že je nalezena více než jedna mezera, jsou všechny následující mezery odstraněny.

Odlišení od návrhu

V této fázi měl být podle návrhu implementován i slovník, oproti kterému by byla získaná slova porovnávána. Byly provedeny praktické experimenty s technologiemi, které automatické opravování pro český jazyk implementují (jmenovitě Python modul `autocorrect`[8] a program `Korektor`[9]). Žádný z nich ovšem nebyl pro účely této práce vhodný z důvodu, že jejich korekce byly občas chybné a nový výsledek byl ještě horší než původní, ze kterého bylo pro člověka možné určit, co bylo původním slovem. Po provedení opravy skrz zmíněné nástroje to však již možné nebylo, což ukazuje tabulka 5.1. Tabulka ilustruje i pokus o zlepšení úspěšnosti převodem na malá písmena. Přestože v některých případech byly výsledky perfektní, to, že jsou i případy, kdy jsou výsledky naopak výrazně zhoršeny, hraje v neprospěch automatického opravování.

Původní text	Výsledek Autocorrectu	Výsledek Korektoru
TVŠ CERNE ZLATO m	VŠ CERN LAT m	TV CERN ZLATO m
tvš cerne zlato m	tv cerne zlato m	tv Herne zlato .
Sovětský flimo wěrnosti	Sovětský fliho věrnosti	Sovětský Klimo věrnosti

Tabulka 5.1: Tabulka ilustrující selhávání automatické korekce textu

I s funkčním slovníkem by však bylo potřeba řešit další problém, a to vlastní jména. S jejich detekováním v textu, který nemusí být sám o sobě korektní, by souvisely další problémy, které by vyžadovaly podstatně komplexnější přístup. Na základě těchto faktů

[8] https://github.com/filyp/autocorrect
[9] https://ufal.mff.cuni.cz/korektor

bylo rozhodnuto, že slovník do této práce implementován nebude, neboť výsledky jsou použitelné i bez něj.

5.6 Nástroj pro import pořadů do databáze

Vkládání pořadů do databáze probíhá na základě čtení textových souborů obsahujících uložené pořady. Implementuje jej nástroj `fillDb.py`. Když je nalezen řádek, který značí úvod hlavičky programu (jeho popisem se zabývá sekce 5.4), jsou tyto informace uloženy a použity pro všechny následující pořady až do doby, než je nalezena hlavička nová. Jak již bylo zmíněno, pro nahrávání do databáze je využita knihovna `SQLAlchemy`. V té jsou vytvořeny třídy reprezentující jednotlivé tabulky databáze, které jsou následně využívány pro určení, do které z tabulek mají být data nahrána.

Detekce jmen a rolí osob

Řádek obsahující pořad je rozdělen do tří částí. První část obsahuje čas vysílání, druhá část název pořadu a třetí část jeho popis. Tyto informace jsou ukládány do samostatných tabulek, aby se zamezilo opakování informací. U popisu dojde dále k pokusu vyhledat osoby, které se na pořadu podílely. K tomu se využívá opět metoda detekce klíčových slov. V případě, že je detekována jedna z hledaných rolí, je jméno osoby za touto rolí uloženo, včetně role samotné, za účelem vložení do databáze. Detekovány jsou role jako např. *Scéna* či *Kamera*. V případě, že má osoba titul zasloužilého umělce či vysokoškolský titul, není tento titul brán v potaz.

Podpora fungování našeptávače

Jména osob, stejně jako názvy pořadů, jsou do tabulek databáze ukládány ve dvou různých formách. První z nich je originální název pořadu či jméno osoby tak, jak bylo detekováno systémem OCR. Druhou formou je uložení stejného jména, tentokrát však převedeného na ASCII charaktery. To je vykonáváno z důvodu zajištění podpory vyhledávání mezi těmito názvy či jmény i bez použití diakritiky.

Nahrávání dat do databáze

Po zpracování všech informací, které řádek s pořadem obsahuje, jsou tyto informace nahrány do databáze. To je ukázáno na kódu, který je možné vidět na výpisu 5.3.

```
1    # vkladani popisu
2    desc_pk = Session.query(Description).filter_by(description=desc).first()
3    if not desc_pk: # popis nebyl v db nalezen, vklada se novy
4        sqlcmd = sqlalchemy.insert(Description).values(description=desc)
5        sqlcmd.compile()
6        insert_res = connection.execute(sqlcmd)
7        desc_pk = insert_res.inserted_primary_key
8        desc_pk = getNormalKey(desc_pk)
9    else: # popis uz v db je, ziskava se jeho PK
10       desc_pk = desc_pk.id
```

Výpis 5.3: Ukázka vkládání popisu pořadu do databáze

V případě nahrávání popisu pořadu je prvně detekováno, zda se takový popis již v databázi nenachází. K tomu je v tomto případě využito vyhledávání v tabulce `Description`,

která popisy obsahuje. V případě, že popis není nalezen, dojde k jeho vložení a získání vloženého primárního klíče. V opačném případě, kdy je tento popis již v databázi přítomen, je získán pouze jeho primární klíč. Obdobný přístup je použit i pro ostatní údaje. Získané primární klíče jsou později vloženy do další tabulky zajišťující, že bude možné správně spojit informace o pořadu.

5.7 Webové uživatelské rozhraní

Jednoduché webové uživatelské rozhraní nad databází bylo vytvořeno, jak bylo již zmíněno, s použitím webového aplikačního rámce Flask. Řídícím nástrojem je `app.py`. Každá stránka, která je v uživatelském rozhraní dostupná, vždy používá pro své fungování šablony a skládá se ze dvou z nich – hlavičky stránky a poté jejího obsahu. Pro nastavení vzhledu stránky je využíván kromě vlastních CSS stylů také rámec *Bootstrap*. Kalendář sloužící pro výběr data a nápovědy při vyhledávání zajišťuje *jQuery*. Pro správné fungování rozhraní je nutný *Javascript*. Jelikož se jedná o uživatelské rozhraní, které nepracuje s žádnými citlivými daty, jsou pro předávání parametrů mezi formuláři a obsluhující aplikací použity pouze metody typu GET.

Hlavní stránka

Ihned po vstupu do uživatelského rozhraní je zobrazen televizní program pro zvolený týden. V případě, že se jedná o první přístup na stránku, nebo datum nebylo nastaveno, je zobrazen program pro 52. týden roku 1992, konkrétně pro 24. 12. 1992. Jelikož je zobrazen vždy televizní program na celý týden, tzn. od pondělí do neděle, je den odpovídající zadanému datu vždy zvýrazněn. Ukázka uživatelského rozhraní, včetně zvýrazněného vybraného dne, je k vidění na obrázku 5.9.

Obrázek 5.9: Ukázka uživatelského rozhraní včetně zvýraznění vyhledávaného data

V uživatelském rozhraní je možné si zvolit stanici, jejíž program chce uživatel vidět. V odlišných letech se tato nabídka liší – například v letech, kdy stanice OK3 nevysílala,

ji není možné v uživatelském rozhraní vybrat, což lze vidět na obrázku 5.10 ilustrujícím výběr stanic, tentokrát ovšem ze dne 13. 12. 1972.

Obrázek 5.10: Ukázka výběru stanice vysílání v případě, kdy bylo zadáno datum 13. 12. 1972

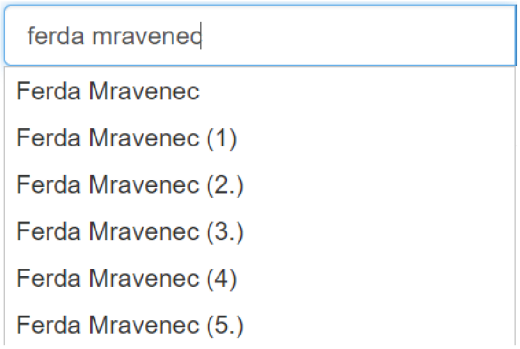
Výběr data

Velmi důležitou součástí uživatelského rozhraní je kalendář sloužící pro výběr data. Ten je lokalizován do českého formátu a umožňuje snadný výběr roku, měsíce a dne. Je spuštěn vždy při stisknutí levého tlačítka myši v oblasti obsahující datum. Je implementováno funkcí `datepicker` z jQuery. V souvislosti s kalendářem jsou ošetřena chybně zadaná data vysílání, takže v případě zadání vyššího či nižšího data, než které se v databázi nachází, dojde v případě vyššího k převedení zadaného data na nejvyšší možné v databázi. Totožně se převede i příliš nízké datum na nejnižší v databázi. Stejný princip je aplikován i na dny v měsíci či měsíce samotné – v případě uvedení neplatné hodnoty dojde k jejímu převedení na krajní hranici, ke které má nejblíž. Když zadaná hodnota formátu data neodpovídá, je nastaveno výchozí datum 24. 12. 1992. Ukázka vzhledu kalendáře je k dispozici na obrázku 5.11.

Obrázek 5.11: Ukázka možností výběru data v kalendáři

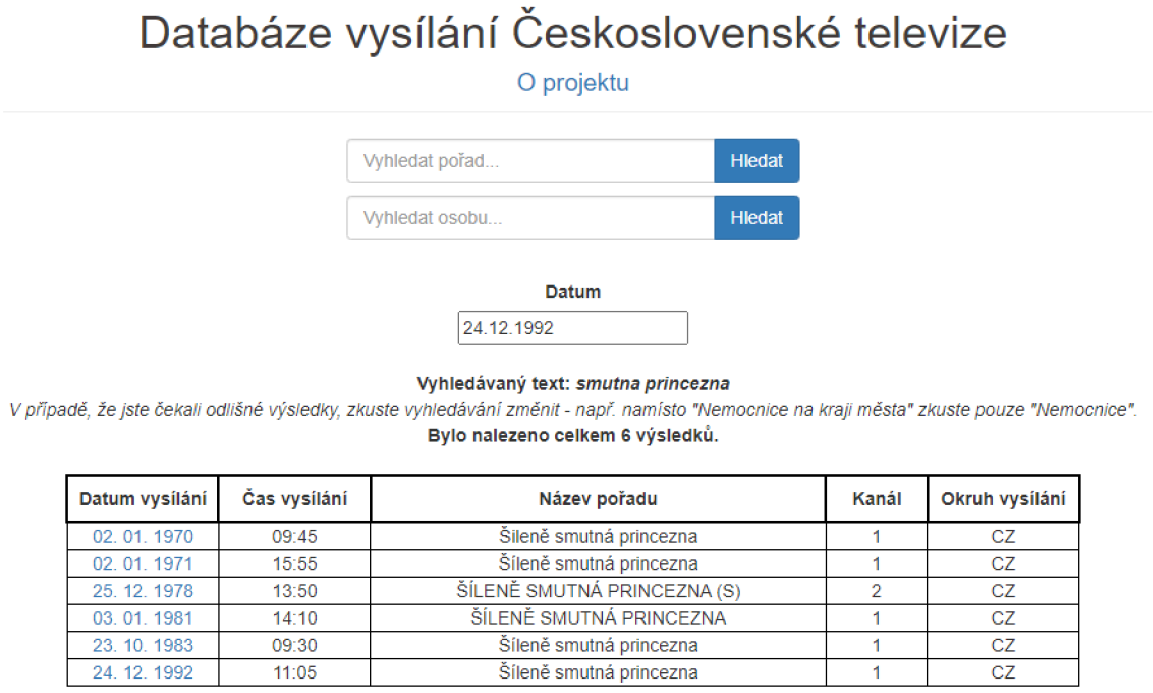
Filtrování pořadů a osob

Další podstatnou součástí rozhraní je možnost vyhledávat pořady a osoby, které se na pořadech podílely. Pro kvalitnější vyhledávání je implementována nápověda, která zobrazí uživateli dostupné pořady či jména na základě zadaného textu již po napsání více než dvou písmen u jmen či tří písmen u názvu pořadů. Pro implementaci nápovědy slouží metoda `autocomplete` pocházející z knihovny jQuery. Po zadání textu je text poslán funkci na pozadí, která zajistí jeho konverzi do ASCII formátu. To umožňuje vyhledávat bez ohledu na diakritiku. Vráceno je patnáct prvních odpovídajících výsledků. Ukázka nápovědy se nachází na obrázku 5.12.



Obrázek 5.12: Nápověda při výběru pořadu

Po vyhledání zadaného výrazu je k dispozici výpis všech výsledků, které byly nalezeny. Současně je k dispozici odkaz na den vysílání konkrétního pořadu a jeho pozici. V případě, že bylo výsledků příliš mnoho (více než 750 u pořadů, u osob není důvod omezovat), je o tom uživatel informován a výsledky nejsou zobrazeny. Tento přístup byl zvolen z důvodu, že vyhledávat v databázi občas desítky tisíc výsledků by mělo za následek prodlevu, která může být až několikavteřinová, což by snižovalo komfort používání. Ukázku výsledků vyhledávání je možné vidět na obrázku 5.13.



Obrázek 5.13: Výsledky vyhledávání mezi pořady po zadání textu *smutna princezna*

.

Kapitola 6

Vyhodnocení úspěšnosti

Tato kapitola se zabývá vyhodnocením úspěšnosti procesu automatické extrakce pořadů z Týdeníku ČS televize. Dojde k vyhodnocení úspěšnosti lokalizace souborů s televizním programem a následně budou vyhodnoceny výsledky úspěšnosti samotné extrakce pořadů. Následovat bude výpis nejčastějších problémů, které se při řešení práce objevily, jejich důvody a nakonec budou představeny možnosti dalšího vývoje.

Úspěšnost detekce televizního programu

V případě detekování souborů obsahujících program byla úspěšnost poměrně vysoká. Docházelo také však k nesprávným detekcím, kdy byla jako televizní program označena část, která jej ve skutečnosti neobsahovala. To se týkalo především devadesátých let, kdy bylo součástí Týdeníku ČS televize například také vysílání rozhlasu, které použitá metoda detekování televizního programu často identifikovala jako televizní programy z důvodu přítomnosti časů vysílání. Pro tuto práci však není takové vysílání relevantní a jedná se o nesprávný výsledek. Z tohoto důvodu a pak také z důvodu, že ne vždy se povedlo nalézt všechny programy, byla nutná manuální kontrola výsledků. Nalezené soubory byly vizuálně ověřeny, že program skutečně obsahují a chybějící soubory byly doplněny.

Úspěšnost byla vyhodnocována jako počet správně identifikovaných programů (*True positive*), se kterými se sečetl počet správně identifikovaných částí, které program neobsahují (*True negative*), což bylo následně vyděleno součtem všech výsledků (tzn. správných i nesprávných). Za nesprávné výsledky jsou považovány soubory s programem, co identifikovány nebyly (*False negative*) a soubory, které byly identifikovány, že program obsahují, ale chybně (*False positive*). Matematický zápis výpočtu, převzatý z [15], je k vidění níže v rovnici 6.1:

$$\text{Přesnost} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6.1}$$

Výsledky automatické detekce televizního programu ve vybraných letech znázorňuje tabulka 6.1. Procentuální úspěšnost byla zaokrouhlena na desetiny procent. Pod pojmem *Program* je v tabulce myšleno *Soubor obsahující televizní program*. Sloupec *Program neobsahuje* značí počet souborů, které byly správně vyhodnoceny, že program neobsahují.

Rok	Programů správně určeno	Programů ve skutečnosti	Programů nesprávně určeno	Program neobsahuje	Souborů celkem	Úspěšnost
1966	95	104	14	307	416	96.6%
1968	47	47	2	327	376	99.5%
1978	52	52	5	368	425	98.8%
1985	208	208	0	212	420	100.0%
1991	712	713	169	943	1824	90.7%

Tabulka 6.1: Úspěšnost detekování programové částí Týdeníku ČS televize ve vybraných letech

Úspěšnost extrakce pořadů

U extrakce pořadů závisela úspěšnost převážně na správné segmentaci strany, což měl na starost použitý systém OCR. V případě, kdy byla tato segmentace nesprávná, musely být výsledky manuálně upravovány. Úspěšnost závisela na formátu televizního programu a v některých letech byla vyšší, v jiných nižší. V pozdějších letech, kdy měl televizní program jednodušší schéma, se úspěšnost automatické segmentace pohybovala okolo *85 %*. U dřívějších let, kdy byl na jedné straně Týdeníku ČS televize program pro více dnů, byly výsledky horší a úspěšnost byla pouze okolo *60 %*. V případě neúspěšné segmentace byly manuálně opravovány stanice, na nichž se daný pořad vysílal, a datum vysílání. Ve výjimečných případech ani manuální korekce nebyla možná a jediným řešením by bylo takovéto dny zcela manuálně přepisovat, což by však bylo v rozporu s cílem této práce, tj. automatickou extrakcí. Takové dny jsou zpracovány pouze částečně.

Samotná úspěšnost extrakce jednotlivých znaků byla po případné manuální korekci chybné segmentace podstatně úspěšnější. Za účelem porovnání výsledků byly manuálně přepsány náhodně vybrané strany televizního programu, které byly posléze porovnány se získanými výsledky, které ale již prošly případnou manuální korekcí segmentace. Kromě toho prošly také korekcí automatickou v podobě úpravy výsledků, která je popsána v kapitole 5.5. Porovnáván byl pouze televizní program 1. programu (nebo jeho pozdější podoby F1). Úspěšnost detekce jednotlivých pořadů a jejich popisu ilustruje tabulka 6.2. Výsledky v tabulce nehledí na úspěšnost oddělení názvu pořadů od jejich popisů, ale na pořady celkově – čas, název a popis pořadů dohromady. Vypočtené hodnoty se odvíjí od počtu znaků, které bylo nutné upravit, odstranit, nebo přidat, aby bylo dosaženo délky manuálně vytvořené, korektní verze textu – je počítána tzv. *Levenshteinova vzdálenost* [16]. Úspěšnost je spočtena jako podobnost textů. Její výpočet ilustruje rovnice 6.2. Výsledky byly zaokrouhleny na desetiny procent.

$$\text{Úspěšnost} = 1 - \frac{\text{Levenshteinova vzdálenost}}{\text{délka delšího textu}} \tag{6.2}$$

Datum	Levenshteinova vzdálenost	Úspěšnost extrakce pořadů
01. 01. 1967	157	94.5 %
04. 01. 1968	215	83.6 %
27. 04. 1970	21	98.6 %
02. 03. 1975	445	83.0 %
08. 05. 1982	6	99.7 %
18. 11. 1990	43	98.6 %

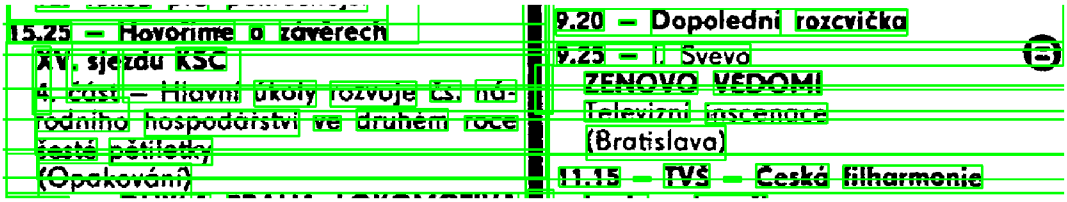
Tabulka 6.2: Úspěšnost extrakce pořadů ve vybraných dnech, včetně jejich popisu

Jak lze z výsledků vyčíst, úspěšnost rozpoznávání pořadů byla relativně vysoká. Obecně lze konstatovat, že v pozdějších letech, kdy měl televizní program jednodušší strukturu, byly výsledky lepší. Jak již bylo ale zmíněno, porovnání těchto výsledků proběhlo až po manuální korekci špatné segmentace stran, bez které by přesnost výsledků v některých případech byla velmi nízká.

6.1 Problémy

Původním záměrem bylo vytvořit databázi vysílání již od roku 1955. Pro to měly být využity naskenované strany z týdeníku *Československý Rozhlas a televize*, což je předchůdce Týdeníku ČS televize. Z důvodu nízké kvality zdrojových materiálů (ukázku lze vidět v příloze C – nejedná se o naskenovaný obraz, ale pouze o fotografii, která není dostatečně ostrá) toto však nebylo možné. Práce se proto zaměřila pouze na roky, které bylo možné zpracovat díky jejich vysoké kvalitě a rozlišení.

Nejzávažnější problém u zpracování Týdeníku ČS televize, kterým bylo selhání automatické segmentace strany do bloků textu, byl již zmíněn. Z nejen tohoto důvodu (druhým důvodem je absence v originálních naskenovaných souborech) nejsou výjimečně některé dny zpracovány, neboť byla segmentace natolik špatná, že nebylo možné program pro daný den získat bez manuálního přepisu celého souboru. Za tento problém je zodpovědný systém OCR Tesseract, který v případě složité grafické úpravy strany není schopen stranu správně automaticky segmentovat. Ukázka části nesprávně segmentované strany se nachází na obrázku 6.1.



Obrázek 6.1: Selhání automatické segmentace, kterou má na starost systémem OCR Tesseract, kdy jsou sloupce interpretovány jako řádky

.

Když se tento problém vyskytl, byl řešen výhradně manuálně. Všechny získané výsledky byly vizuálně zkontrolovány a v případě, že se u nich segmentace nepodařila, jak by měla, byly rozděleny do původních bloků manuálně a případné neodpovídající pořady byly odstraněny. Příklad takové korekce lze vidět na výpisu níže. Výpis 6.1 obsahuje zkrácený výsledek s nepovedenou automatickou segmentací. Výpis 6.2 ilustruje manuální korekci.

```
CZ
28. 2. 1991
1
0.10 Zprávy NAZEVPORADU
23.35 OK 3 plus NAZEVPORADU
0.05 Závěr vysílání NAZEVPORADU
0.15 Závěr vysílání NAZEVPORADU
```

Výpis 6.1: Neúspěšná automatická segmentace pořadů

```
CZ
28. 2. 1991
2
23.35 OK 3 plus NAZEVPORADU
0.05 Závěr vysílání NAZEVPORADU

CZ
28. 2. 1991
1
0.10 Zprávy NAZEVPORADU
0.15 Závěr vysílání NAZEVPORADU
```

Výpis 6.2: Manuálně opravená segmentace

Kromě segmentace byly nejčastějšími problémy nesprávně klasifikované znaky. To bylo způsobeno například tiskařskými chybami v textu či nedokonalostí použitého systému OCR. Případ, kdy byla již v originálním textu přítomna chyba, je možné vidět na obrázku 6.2. Správně by měl čas být `12.00`, systém OCR jej však ze zmíněného důvodu detekuje jako `42.00`.

Obrázek 6.2: Tiskařská chyba v názvu pořadu – správně má být 12.00

.

6.2 Další vývoj

V rámci dalšího vývoje či zlepšení databáze by bylo možné využít speciální nástroj pro zpracování přirozeného jazyka za účelem detekce rolí a jmen osob podílejících se na jednotlivých pořadech. Současné řešení je sice funkční, ale není perfektní například v případě, že má osoba neobvyklý titul před jménem.

Vyhledané osoby v jednotlivých pořadech a nebo také pořady samotné by mohly být v rámci dalšího vývoje spojeny s dalšími informacemi o nich. Konkrétně by v případě domluvy s vlastníky mohlo dojít k propojení s Česko-Slovenskou filmovou databází[1]. V ideálním případě by se přes jméno osoby bylo možné hypertextovým odkazem přesunout na její profil na zmíněné stránce. S využitím stejného principu by mohly na tuto stránku vést i názvy pořadů.

Tato práce se zabývá vysíláním Československé Televize. Po rozpadu Československa ale televizní vysílání pokračovalo. Databáze by proto mohla být rozšířena o další roky až do současnosti, nebo naopak být navázána na jiné existující řešení. V případě, že budou získány dostatečně kvalitní obrazy, by mohlo dojít také k rozšíření o vysílání před rokem 1966.

[1] https://www.csfd.cz/

Kapitola 7

Závěr

Cílem této práce bylo vytvořit databázi vysílání Československé televize s využitím zdrojových materiálů v podobě naskenovaných stran Týdeníku ČS televize. Podařilo se vytvořit téměř kompletní databázi vysílání pro roky 1966 až 1992. Ve výsledcích je přítomno malé množství chyb, které jsou způsobeny použitým systémem OCR, kterým byl Tesseract.

V práci byly popsány základy zpracování textu s využitím technologie OCR a volně dostupné nástroje, které OCR využívají. Následně byly představeny relační databáze včetně procesu jejich normalizace a existujících databázových systémů. Následovala analýza požadavků a dat, poté obecný návrh řešení a nakonec jeho konkrétní implementace a zhodnocení výsledků.

Vytvořená databáze je přístupná přes jednoduché webové uživatelské rozhraní. To nabízí možnost prohlížet televizní pořady vysílané v jednotlivých dnech. Dále nabízí uživateli možnost mezi vysílanými pořady vyhledávat, což je usnadněno přítomností našeptávání výsledků na základě zadaného vstupu. Vyhledávat je možno také mezi osobami, které se na pořadech podílely.

Při extrakci pořadů z jednotlivých souborů občas docházelo k problémům spočívajícím v neúspěšné automatické segmentaci televizního programu na části, což muselo být manuálně korigováno. Po této korekci je ovšem možno konstatovat, že výsledky byly v drtivé většině zpracovaných dnů dostatečně kvalitní s úspěšností pohybující se okolo 90 % a tvorbu databáze lze označit za úspěšnou.

Je mnoho možností, jak práci dále vyvíjet. Mezi hlavní z nich patří rozšíření databáze o vysílání před rokem 1965, pro které v době psaní práce nebyly k dispozici dostatečně kvalitní naskenované zdrojové materiály. Společně s tím by databáze mohla být obohacena o vysílání po roce 1992. Další možností by mohlo být propojení s Česko-Slovenskou filmovou databází.

Literatura

[1] Adamus, J. a Hošek, J. Program 68. *Česká televize* [online]. 2018 [cit. 2022-04-14]. Dostupné z: https://www.ceskatelevize.cz/specialy/totostoleti/program-68.

[2] Chaudhuri, A., Mandaviya, K., Badelia, P. a Ghosh, S. K. *Optical Character Recognition Systems for Different Languages with Soft Computing*. 1. vyd. Springer International Publishing, 2017. ISBN 978-3-319-50252-6.

[3] Cheriet, M., Kharma, N., Liu, C.-L. a Suen, C. Y. *Character Recognition Systems: A Guide for Students and Practioners*. 1. vyd. John Wiley & Sons, 2007. ISBN 978-0-471-41570-1.

[4] Dilmegani, C. Best OCR by Text Extraction Accuracy in 2022. *AIMultiple* [online]. 2021. Revidováno 11. 2. 2022 [cit. 2022-04-21]. Dostupné z: https://research.aimultiple.com/ocr-accuracy.

[5] Guruprasad, P., Mahalingpur, K. S. a Manjesh, T. N. Overview of different thresholding methods in image processing. In: *TEQIP Sponsored 3rd National Conference on ETACC* [online]. červen 2020 [cit. 2022-05-01]. Dostupné z: https://www.researchgate.net/publication/342038946.

[6] Karthick, K., Ravindrakumar, K., Francis, R. a Ilankannan, S. Steps Involved in Text Recognition and Recent Research in OCR; A Study. *International Journal of Recent Technology and Engineering*. Květen 2019, sv. 8, s. 3095–3100. ISSN 2277-3878.

[7] Li, S. Z. a Jain, A. Local Adaptive Thresholding. In: *Encyclopedia of Biometrics* [online]. Boston, MA: Springer US, 2009, s. 939. DOI: 10.1007/978-0-387-73003-5_506. ISBN 978-0-387-73003-5. Dostupné z: https://doi.org/10.1007/978-0-387-73003-5_506.

[8] MariaDB Foundation. About MariaDB Server. *MariaDB Foundation - MariaDB.org* [online]. 2022 [cit. 2022-04-11]. Dostupné z: https://mariadb.org/about/.

[9] Oppel, A. J. *Databases: A Beginner's Guide*. 1. vyd. The McGraw-Hill Companies, 2009. ISBN 978-0-07-160847-3.

[10] Oracle Corporation. What is MySQL. *MySQL :: Developer Zone* [online]. 2022 [cit. 2022-04-11]. Dostupné z: https://dev.mysql.com/doc/refman/8.0/en/what-is-mysql.html.

[11] Římský, P. *Historie Redakce sportu České televize v letech 1993-2015*. 2016. Diplomová práce. Univerzita Karlova, Fakulta sociálních věd, Katedra mediálních studií. Vedoucí práce Bednařík, P.

[12] Saba, T., Sulong, G. a Rehman, A. A Survey on Methods and Strategies on Touched Characters Segmentation. *International Journal of Research and Reviews in Computer Science*. Leden 2010, sv. 1, s. 103–114. ISSN 2079-2557.

[13] Solid IT. DB-Engines Ranking of Relational DBMS. *DB-Engines - Knowledge Base of Relational and NoSQL Database Management Systems* [online]. Duben 2022 [cit. 2022-04-11]. Dostupné z: https://db-engines.com/en/ranking/relational+dbms.

[14] SQLite. About SQLite. *SQLite Home Page* [online]. [cit. 2022-04-11]. Dostupné z: https://www.sqlite.org/about.html.

[15] Tharwat, A. Classification assessment methods. *Applied Computing and Informatics*. 2021, sv. 17, č. 1, s. 168–192. ISSN 2634-1964.

[16] Wu, G. String Similarity Metrics – Edit Distance. *Baeldung on CS* [online]. 2020. Revidováno 5. 11. 2021 [cit. 2022-05-02]. Dostupné z: https://www.baeldung.com/cs/string-similarity-edit-distance.

[17] Česká televize. Prehistorie – Československá televize do roku 1992. *Česká televize* [online]. [cit. 2022-04-11]. Dostupné z: https://www.ceskatelevize.cz/vse-o-ct/historie/ceskoslovenska-televize/prehistorie.

[18] Šimůnek, M. *SQL: Kompletní kapesní průvodce*. 1. vyd. GRADA Publishing, 1999. ISBN 80-7169-692-7.

Příloha A

Obsah přiloženého paměťového média

- **doc/** – Zdrojové kódy písemné zprávy v LaTeXu
- **README.txt** – Popis obsahu média, návod ke spuštění a další informace
- **resources/** – Knihovny a nástroje potřebné pro spuštění
- **src/** – Zdrojové kódy programů, databáze a další soubory
- **tv_examples/** – Ukázky televizního programu
- **xfiala61-databaze-vysilani.pdf** – Písemná zpráva
- **xfiala61-databaze-vysilani-tisk.pdf** – Písemná zpráva bez barevných odkazů

Příloha B

Ukázky televizního programu z Týdeníku ČS televize

Pořady Čs. televize a výběr z programů Čs. rozhlasu v týdnu od 27. prosince 1965–2. ledna 1966

Televize

PONDĚLÍ

27. prosince 1965

ČTVRTEK

30. prosince 1965

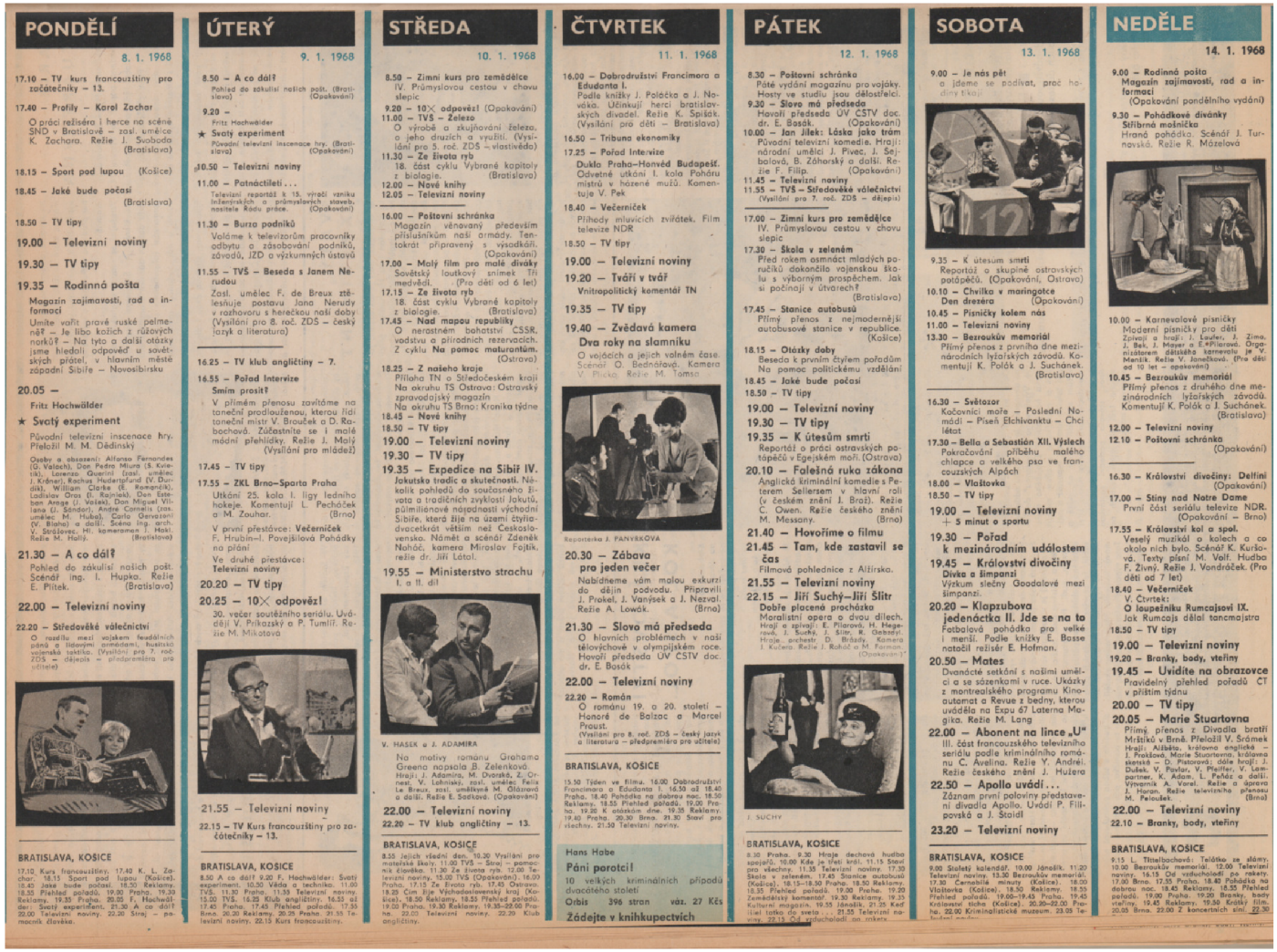
ÚTERÝ

28. prosince 1965

STŘEDA

29. prosince 1965

O SPENGLERŮV POHÁR



PONDĚLÍ

8. 1. 1968

17.10 — TV kurs francouzštiny pro začátečníky — 13.

17.40 — Profily — Karol Zachar

O práci režiséra i herce na scéně SND v Bratislavě — zasl. umělce K. Zachara. Režie J. Svoboda (Bratislava)

18.15 — Sport pod lupou (Košice)

18.45 — Jaké bude počasí (Bratislava)

18.50 — TV tipy

19.00 — Televizní noviny

19.30 — TV tipy

19.35 — Rodinná pošta

Magazín zajímavostí, rad a informací

Umíte vařit pravé ruské pelmeně? — Je líbo kožich z růžových norků? — Na tyto a další otázky jsme hledali odpověď u sovětských přátel, v hlavním městě západní Sibiře — Novosibirsku

20.05 —

Fritz Hochwälder

★ Svatý experiment

Původní televizní inscenace hry. Přeložil M. M. Dědinský

Osoby a obsazení: Alfonso Fernandes (G. Valach), Don Pedro Miura (S. Kvietik), Lorenzo Querini (zasl. umělec J. Kröner), Rachus Hedertpfund (V. Durdík), William Clarke (E. Romančík), Ladislav Oros (I. Rajnichl), Don Esteban Arnge (J. Valek), Don Miguel Villana (J. Sándor), André Corneis (zasl. umělec M. Huba), Carlo Gervaroni (V. Blaha) a další. Scéna Ing. arch. V. Strášlove, Hl. kameramen J. Hoki. Režie M. Holý. (Bratislava)

21.30 — A co dál?

Pohled do zákulisí našich pošt. Scénář Ing. I. Hupka. Režie E. Přítek. (Bratislava)

22.00 — Televizní noviny

22.20 — Středověké válečnictví

O rozdílu mezi vojskem feudálních pánů a lidovými armádami, husitská vojenská taktika. (Vysílání pro 7. roč. ZDŠ — dějepis — předpremiéra pro učitele)

BRATISLAVA, KOŠICE

17.10 Kurs francouzštiny. 17.40 K. L. Zachar. 18.15 Sport pod lupou (Košice). 18.45 Jaké bude počasí. 18.50 Reklamy. 18.55 Přehled pořadů. 19.00 Praha. 19.30 Reklamy. 19.35 Praha. 20.05 F. Hochwälder: Svatý experiment. 21.30 A co dál? 22.00 Televizní noviny. 22.20 Stroj — pomocník člověka.

ÚTERÝ

9. 1. 1968

8.50 — A co dál?

Pohled do zákulisí našich pošt. (Bratislava) (Opakování)

9.20 —

Fritz Hochwälder

★ Svatý experiment

Původní televizní inscenace hry. (Bratislava) (Opakování)

10.50 — Televizní noviny

11.00 — Patnáctiletí . . .

Televizní reportáž k 15. výročí vzniku inženýrských a průmyslových staveb, nositele Řádu práce. (Opakování)

11.30 — Burza podniků

Voláme k televizorům pracovníky odbytu a zásobování podniků, závodů, JZD a výkupních ústavů

11.55 — TVŠ — Beseda s Janem Nerudou

Zasl. umělec F. de Breux ztělesňuje postavu Jana Nerudy v rozhovoru s hereckou naší doby (Vysílání pro 8. roč. ZDŠ — český jazyk a literatura)

16.25 — TV klub angličtiny — 7.

16.55 — Pořad Intervize

Smím prosit?

V přímém přenosu zavítáme na taneční prodloužence, kterou řídí taneční mistr V. Brouček a D. Rabochová. Zúčastníte se i malé módní přehlídky. Režie J. Malý (Vysílání pro mládež)

17.45 — TV tipy

17.55 — ZKL Brno—Sparta Praha

Utkání 25. kola I. ligy ledního hokeje. Komentují L. Pecháček a M. Zouhar. (Brno)

V první přestávce: Večerníček F. Hrubín—I. Povejšilová Pohádky na přání

Ve druhé přestávce: Televizní noviny

20.20 — TV tipy

20.25 — 10× odpovězi!

30. večer soutěžního seriálu. Uvádějí V. Přikaszký a P. Tumlíř. Režie M. Mikotová

21.55 — Televizní noviny

22.15 — TV Kurs francouzštiny pro začátečníky — 13.

BRATISLAVA, KOŠICE

8.50 A co dál? 9.20 F. Hochwälder: Svatý experiment. 10.50 Věda a technika. 11.00 TVŠ. 11.30 Praha. 11.55 Televizní noviny. 15.00 TVŠ. 16.25 Klub angličtiny. 16.55 až 17.45 Praha. 17.45 Přehled pořadů. 17.55 Brno. 20.20 Reklamy. 20.25 Praha. 21.55 Televizní noviny. 22.15 Kurs francouzštiny.

STŘEDA

10. 1. 1968

8.50 — Zimní kurs pro zemědělce

IV. Průmyslovou cestou v chovu slepic

9.20 — 10× odpovězi! (Opakování)

11.00 — TVŠ — Železo

O výrobě a zkušťování železa, o jeho druzích a využití. (Vysílání pro 5. roč. ZDŠ — vlastivěda)

11.30 — Ze života ryb

18. část cyklu Vybrané kapitoly z biologie. (Bratislava)

12.00 — Nové knihy

12.05 — Televizní noviny

16.00 — Poštovní schránka

Magazín věnovaný především příslušníkům naší armády. Tentokrát připravený s vysvádkáři. (Opakování)

17.00 — Malý film pro malé diváky

Sovětský loutkový snímek Tři medvědi. (Pro děti od 6 let)

17.15 — Ze života ryb

18. část cyklu Vybrané kapitoly z biologie. (Bratislava)

17.45 — Nad mapou republiky

O nerostném bohatství ČSSR, vodstvu a přírodních rezervacích. Z cyklu Na pomoc maturantům. (Ostrava)

18.25 — Z našeho kraje

Příloha TN o Středočeském kraji Na okruhu TS Ostrava: Ostravský zpravodajský magazín

Na okruhu TS Brno: Kronika týdne

18.45 — Nové knihy

18.50 — TV tipy

19.00 — Televizní noviny

19.30 — TV tipy

19.35 — Expedice na Sibiř IV.

Jakutsko tradic a skutečnosti. Několik pohledů do současného života a tradičních zvyklostí Jakutů, půlmiliónové národnosti východní Sibiře, která žije na území čtyřiadvacetkrát větším než Československo. Námět a scénář Zdeněk Noháč, kamera Miroslav Fojtík, režie dr. Jiří Látal.

19.55 — Ministerstvo strachu

I. a II. díl

V. HÁSEK a J. ADAMIRA

Na motivy románu Grahama Greena napsala B. Zelenková. Hrají: J. Adamira, M. Dvorská, Z. Ornest, V. Lohniský, zasl. umělec Felix Le Breux, zasl. umělkyně M. Glázrová a další. Režie E. Sadková. (Opakování)

22.00 — Televizní noviny

22.20 — TV klub angličtiny — 13.

BRATISLAVA, KOŠICE

8.50 Jejich všední den. 10.30 Vysílání pro mateřské školy. 11.00 TVŠ — Stroj — pomocník člověka. 11.30 Ze života ryb. 12.00 Televizní noviny. 15.00 TVŠ (Opakování). 16.00 Praha. 17.15 Ze života ryb. 17.45 Ostrava. 18.25 Cim žije Východoslovenský kraj (Košice). 18.50 Reklamy. 18.55 Přehled pořadů. 19.00 Praha. 19.30 Reklamy. 19.35—22.00 Praha. 22.00 Televizní noviny. 22.20 Klub angličtiny.

ČTVRTEK

11. 1. 1968

16.00 — Dobrodružství Francimora a Eduganta I.

Podle knížky J. Poláčka a J. Nováka. Účinkují herci bratislavských divadel. Režie K. Spišák. (Vysílání pro děti — Bratislava)

16.50 — Tribuna ekonomiky

17.25 — Pořad Intervize

Dukla Praha—Honvéd Budapešť. Odvetné utkání I. kola Poháru mistrů v házené mužů. Komentuje V. Pek

18.40 — Večerníček

Příhody mluvících zvířátek. Film televize NDR

18.50 — TV tipy

19.00 — Televizní noviny

19.20 — Tváří v tvář

Vnitropolitický komentář TN

19.35 — TV tipy

19.40 — Zvědavá kamera

Dva roky na slamníku

O vojácích a jejich volném čase. Scenár O. Bednářová. Kamera V. Plicka. Režie M. Tomsa

Reportérka J. PANYRKOVÁ

20.30 — Zábava pro jeden večer

Nabídneme vám malou exkurzi do dějin podvodu. Připravili J. Prokel, J. Vanýsek a J. Nezval. Režie A. Lowák. (Brno)

21.30 — Slovo má předseda

O hlavních problémech v naší tělovýchově v olympijském roce. Hovoří předseda ÚV ČSTV doc. dr. E. Basák

22.00 — Televizní noviny

22.20 — Román

O románu 19. a 20. století — Honoré de Balzac a Marcel Proust.

(Vysílání pro 8. roč. ZDŠ — český jazyk a literatura — předpremiéra pro učitele)

BRATISLAVA, KOŠICE

13.50 Týden ve filmu. 16.00 Dobrodružství Francimora a Eduganta I. 16.50 až 18.40 Praha. 18.40 Pohádka na dobrou noc. 18.50 Reklamy. 18.55 Přehled pořadů. 19.00 Praha. 19.20 K otázkám dne. 19.35 Reklamy. 19.40 Praha. 20.30 Brno. 21.30 Stavě pro všechny. 21.50 Televizní noviny.

PÁTEK

12. 1. 1968

8.30 — Poštovní schránka

Páté vydání magazínu pro vojáky. Hosty ve studiu jsou dělostřelci.

9.30 — Slovo má předseda

Hovoří předseda ÚV ČSTV doc. dr. E. Basák. (Opakování)

10.00 — Jan Jílek: Láska jako trám

Původní televizní komedie. Hrají: národní umělci J. Pivec, J. Šejbalová, B. Záhorský a další. Režie F. Filip. (Opakování)

11.45 — Televizní noviny

11.55 — TVŠ — Středověké válečnictví

(Vysílání pro 7. roč. ZDŠ — dějepis)

17.00 — Zimní kurs pro zemědělce

IV. Průmyslovou cestou v chovu slepic

17.30 — Škola v zeleném

Před rokem osmnáct mladých poručíků dokončilo vojenskou školu s výborným prospěchem. Jak si počínají v útvarech?

(Bratislava)

17.45 — Stanice autobusů

Přímý přenos z nejmodernější autobusové stanice v republice. (Košice)

18.15 — Otázky doby

Beseda k prvním čtyřem pořadům Na pomoc politickému vzdělání

18.45 — Jaké bude počasí

18.50 — TV tipy

19.00 — Televizní noviny

19.30 — TV tipy

19.35 — K útesům smrti

Reportáž o práci ostravských potápěčů v Egejském moři. (Ostrava)

20.10 — Falešná ruka zákona

Anglická kriminální komedie s Peterem Sellersem v hlavní roli (v českém znění Ladislav Boháč). Režie C. Owen. Režie českého znění M. Messany. (Brno)

21.40 — Hovoříme o filmu

21.45 — Tam, kde zastavil se čas

Filmová pohlednice z Altýrska.

21.55 — Televizní noviny

22.15 — Jiří Suchý—Jiří Slitr

Dobře placená procházka

Moralistní opera o dvou dílech. Hrají a zpívají: E. Pilarová, H. Hegerová, J. Suchý, J. Slitr, K. Gabrielová. Hraje orchestr D. Brázdy. Kamera J. Kučera. Režie J. Roháč a M. Forman. (Opakování)

J. SUCHÝ

BRATISLAVA, KOŠICE

8.30 Praha. 9.30 Hraje dechová hudba spojařů. 10.00 Kde je třetí král. 11.15 Slovo pro všechny. 11.35 Televizní noviny. 17.30 Škola v zeleném. 17.45 Stanice autobusů (Košice). 18.15—18.50 Praha. 18.50 Reklamy. 18.55 Přehled pořadů. 19.00 Praha. 19.20 Zemědělský komentář. 19.30 Reklamy. 19.35 Kulturní magazín. 19.55 Jánošík. 21.25 Kdo šílel tatka do světa . . . 21.55 Televizní noviny. 22.15 Od vrabčoholdí po rakety.

SOBOTA

13. 1. 1968

9.00 — Je nás pět

a jdeme se podívat, proč hodiny tikají

9.35 — K útesům smrti

Reportáž o skupině ostravských potápěčů. (Opakování, Ostrava)

10.10 — Chvilka v maringotce

Den dřevára (Opakování)

10.45 — Písničky kolem nás

11.00 — Televizní noviny

13.30 — Bezroukův memoriál

Přímý přenos z prvního dne mezinárodních lyžařských závodů. Komentují K. Polák a J. Suchánek. (Bratislava)

16.30 — Světozor

Kočovníci moře — Poslední Nomádi — Píseň Elchivanktu — Chci létat

17.30 — Bella a Sebastián XII. Výsledek

Pokračování příběhu malého chlapce a velkého psa ve francouzských Alpách

18.00 — Vlaštovka

18.50 — TV tipy

19.00 — Televizní noviny + 5 minut o sportu

19.30 — Pořad k mezinárodním událostem

19.45 — Království divočiny

Dívka a šimpanzi

Výzkum slečny Goodallové mezi šimpanzi.

20.20 — Klapzubova jedenáctka II. Jde se na to

Fotbalová pohádka pro velké i menší. Podle knížky E. Basse natočil režisér E. Hofman.

20.50 — Mates

Dvanácté setkání s našimi umělci a se sázenkami v ruce. Ukázky z montrealského programu Kinoautomat a Revue z bedny, kterou uváděla na Expu 67 Laterna Magika. Režie M. Lang

22.00 — Abonent na lince „U"

III. část francouzského televizního seriálu podle kriminálního románu C. Avelina. Režie Y. Andréi. Režie českého znění J. Hužera

22.50 — Apollo uvádí . . .

Záznam první poloviny představení divadla Apollo. Uvádí P. Filipovská a J. Štaidl

23.20 — Televizní noviny

BRATISLAVA, KOŠICE

9.00 Století kalendář. 10.00 Jánošík. 11.20 Televizní noviny. 13.30 Bezroukův memoriál. 17.30 Čarnobílé minuty (Košice). 18.00 Vlaštovka (Košice). 18.50 Reklamy. 18.55 Přehled pořadů. 19.00—19.45 Praha. 19.45 Království ticha (Košice). 20.20—22.00 Praha. 22.00 Kriminalistické muzeum. 23.05 Televizní noviny.

NEDĚLE

14. 1. 1968

9.00 — Rodinná pošta

Magazín zajímavostí, rad a informací (Opakování pondělního vydání)

9.30 — Pohádkové divánky

Stříbrná mošnička

Hraná pohádka. Scénář J. Turnavská. Režie R. Mázrelová

10.00 — Karnevalové písničky

Moderní písničky pro děti Zpívají a hrají: J. Lauer, J. Zima, J. Bek, J. Mayer a E.Pilarová. Organizátorem dětského karnevalu je V. Mantík. Režie V. Janečková. (Pro děti od 10 let — opakování)

10.45 — Bezroukův memoriál

Přímý přenos z druhého dne mezinárodních lyžařských závodů. Komentují K. Polák a J. Suchánek. (Bratislava)

12.00 — Televizní noviny

12.10 — Poštovní schránka (Opakování)

16.30 — Království divočiny: Delfíni (Opakování)

17.00 — Stíny nad Notre Dame

První část seriálu televize NDR. (Opakování — Brno)

17.55 — Království kol a spol.

Veselý muzikál o kolech a co okolo nich bylo. Scénář K. Kuršová. Texty písní M. Volf. Hudba F. Živný. Režie J. Vandráček. (Pro děti od 7 let)

18.40 — Večerníček

V. Čtvrtek:

O loupežníku Rumcajsovi IX.

Jak Rumcajs dělal tancmajstra

18.50 — TV tipy

19.00 — Televizní noviny

19.20 — Branky, body, vteřiny

19.45 — Uvidíte na obrazovce

Pravidelný přehled pořadů ČT v příštím týdnu

20.00 — TV tipy

20.05 — Marie Stuartovna

Přímý přenos z Divadla bratří Mrštíků v Brně. Přeložil V. Šrámek Hrají: Alžběta, královna anglická — I. Prokšová, Marie Stuartovna, královna skotská — D. Pistorová; dále hrají: J. Dubek, V. Pivec, V. Pfeiffer, V. Lampartner, K. Adam, L. Peňáz a další. Výtvarník A. Varel. Režie a aparát J. Horan. Režie televizního přenosu M. Peloušek. (Brno)

22.00 — Televizní noviny

22.10 — Branky, body, vteřiny

BRATISLAVA, KOŠICE

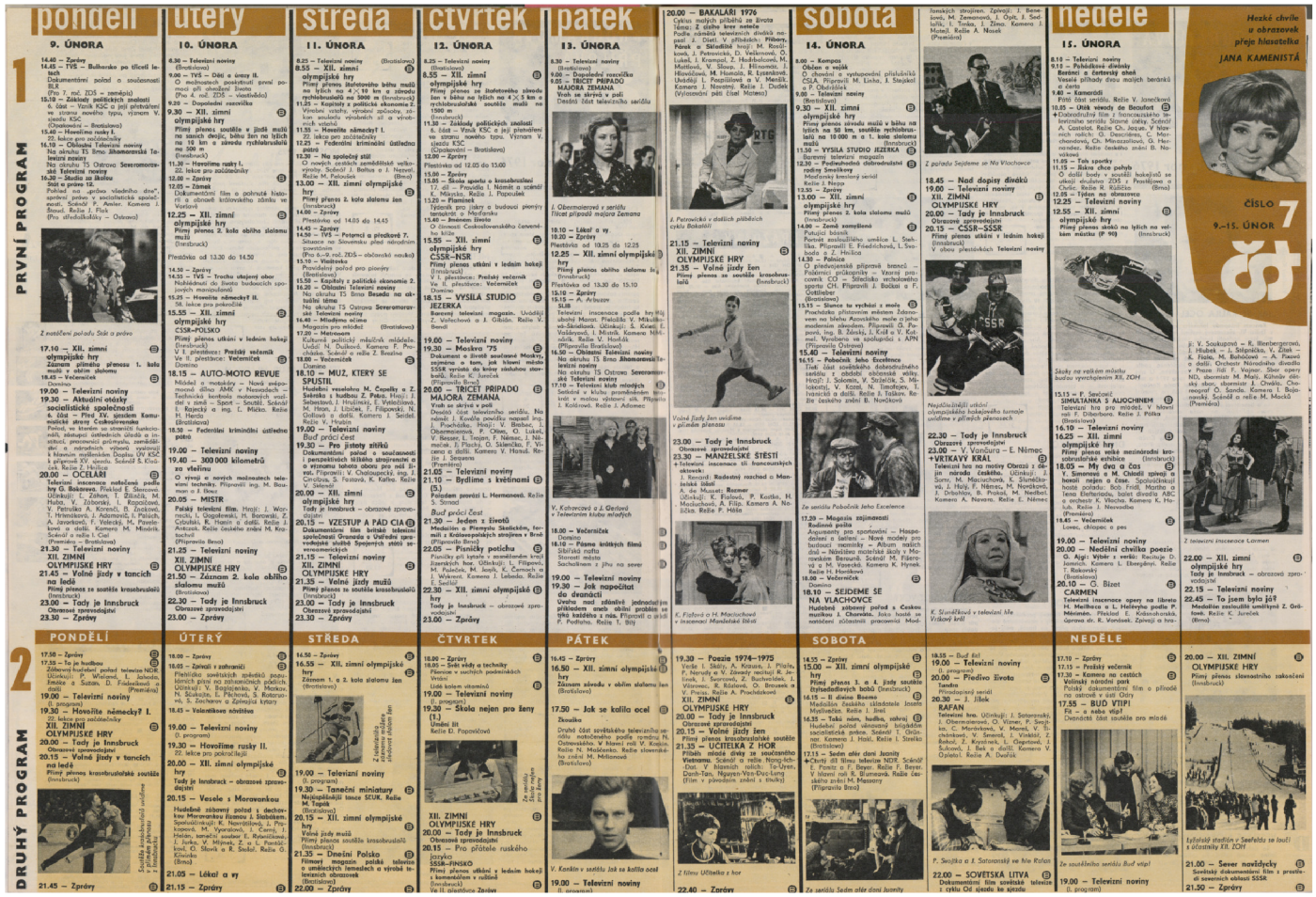
9.15 L. Titatšachová: Telátko ze slámy. 10.00 Bezroukův memoriál. 12.00 Televizní noviny. 16.15 Od vzduchoholdí po rakety. 17.00 Brno. 17.55 Praha. 18.40 Pohádka na dobrou noc. 18.45 Reklamy. 18.55 Přehled pořadů. 19.00 Praha. 19.20 Branky, body vteřiny. 19.45 Reklamy. 19.50 Krátký film. 20.05 Brno. 22.00 Z koncertních sín. 22.30

Obrázek B.2: Formát programu pro roky 1968-1970

Obrázek B.3: Formát programu pro 2. část roku 1970

Obrázek B.4: Formát programu pro roky 1970-1971

Obrázek B.5: Formát programu pro roky 1972-1975



Obrázek B.6: Formát programu pro roky 1976-1978

středa 30. prosince

(1)

8.55 — Zprávy
9.00 — Rohatá princezna
Krásná pohádka podle J. Š. Kubína
9.15 — M. Mathausová
SLUNÍČKO NA HOUPAČCE
Televizní hra

M. Mikulášová a O. Sklenčka ve hře Sluníčko na houpačce

10.40 — Knihy pro vás
10.45 — Buď — anebo
Zábavný soutěžní seriál. Kamera I. Koudelka. Režie J. Vondráček
11.25 — Zprávy
Přestávka od 11.40 do 16.25
16.25 — Zprávy
16.30 — Spokojené stáří
O podmínkách spokojeného soužití ve třetím období života. Připravili D. Morunková a P. Obdržálek
17.15 — Příležitost pro talenty
O zájmové umělecké činnosti mladých, tentokrát o diskotékách. Scénář J. Dolák. Režie R. Benko (Bratislava)
17.55 — Klaun Ferdinand (8.)
Seriál televize NDR pro nejmenší. V hlavní roli J. Vršťala. Režie T. Meinl. Režie českého znění S. Sázavská (Připravilo Brno)
18.25 — Knihy pro vás
18.30 — Na poslední chvíli ●
O hrdinském činu příslušníka tankového vojska. Připravil J. Valchář
18.45 — Z galérie hrdinů
Ján Bicháč
Profil hrdiny socialistické práce. Scénář F. Hrubý. Režie J. Urblík (Bratislava)
19.10 — Večeníček
Maxipes Fík

19.30 — Televizní noviny
20.00 — J. Dietl
NEMOCNICE NA KRAJI MĚSTA (20.)
Závěr televizního seriálu. Hrají: L. Chudík, J. Abrhám, L. Frej, M. Kopecký, E. Balzerová, H. Maciuchová, J. Dvořák, I. Janžurová, J. Štěpánková, O. Kaiser, F. Němec, J. Henzlik, A. Čunderlíková, L. Pešek, D. Medřická, V. Sloup, N. Popelíková a další. Kamera J. Novotný. Režie J. Dudek (Premiéra)

E. Balzerová a I. Abrhám v seriálu Nemocnice na kraji města

21.00 — Z melodie do melodie aneb Revue na střeše
Účinkují: Orchestr a balet Čs. televize, dirigent M. Krebs, choreografie I. Kubicová, H. Vondráčková, J. Korn, E. Pilarová, V. Špinarová, K. Hála, J. Malásek, M. Tučný, P. Bobek, K. Černoch a další. Kamera J. Osten. Scénář a režie J. Bonaventura (Premiéra)

H. Vondráčková v pořadu Z melodie do melodie aneb Revue na střeše

21.45 — 24 hodin ve světě
22.00 — ČESKOSLOVENSKO 1981
Přehled vnitropolitických událostí
22.40 — Zprávy

BRATISLAVA, KOŠICE
■ 8.55 Zprávy. 9.00 Housle. 9.15 Praha. 10.40 Knihy pro vás (Košice). 10.45 Praha. 11.25 Hovoříme francouzsky. 11.55 Zprávy. 15.55 Zprávy. 16.30 Hovoříte německy? 16.50 Tři mušketýři. 17.15 Příležitost pro talenty. 17.55 Klaun Ferdinand. 18.25 Knihy týdne (Košice). 18.30 Praha. 18.45 Z galérie hrdinů. 19.10 Večeníček. 19.30 až 22.40 Praha. 22.40 Zprávy.

(2)

15.55 — Zprávy
16.00 — Pohár osvobození ●
Přímý přenos části finálového utkání mezinárodního dorosteneckého turnaje v ledním hokeji (Brno)
17.25 — Branná hlídka
17.30 — O autech a lidech (6.)
Závěr francouzského televizního seriálu. Připravili H. de Turenne a A. Barret (Premiéra — Bratislava)
18.25 — Máme šance ●
Závěrečná část soutěže TKM pro mladé dvojice. Režie E. Sakolovský ml. (Připravilo Brno)
19.15 — Lékař a vy
19.25 — Zkus rytmus (Bratislava)
19.30 — Středočeská galérie
Připravili E. Friedrichová, P. Volf a B. Šefranka
20.00 — Indické klasické a lidové tance
Indický hudební film (Premiéra)
21.00 — Kouzelná světlna
Poetický film o předchůdci filmu — laterně magice
21.15 — Aktuality
21.45 — PROROK, ZLATO A TRANSYLVÁNCI
Rumunský dobrodružný film. Scénář T. Popovici. Režie D. Pita. Hrají: I. Ciobanu, O. I. Moldoveanu, M. Diaconu, V. Rebegiuc, L. Tudorachová, C. Bertalová a další (Televizní premiéra — v původním znění s titulky)

Z filmu Prorok, zlato a Transylvánci

BRATISLAVA, KOŠICE
■ 16.40 Zprávy. 16.55 Z galérie vynikajících sportovců. 17.30 O autech a lidech. 18.25 Praha. 19.15 Z deníku VB 55R. 19.25 Zkus rytmus. 19.30 V Ružomberku pod střechou. 20.00 Operetní přátelství. 20.55 Mistrovská díla Drážďanské galérie. 21.15 Aktuality. 21.45 Prorok, zlato a Transylvánci.

● — černobílý pořad

(1) 16.30

SPOKOJENÉ STÁŘÍ

Kdy je člověk starý? Stáří je materiální = kalendářní věk, biologické = jak organismus funguje, jak pracuje, konvenční = kdy je kdo považován za starého, úřední, administrativní = norma zákona, důchod. Člověk je tak starý, jak se cítí a co ve svém vlastním zájmu ví o stáří on a také ti, s nimiž žije. V pořadu se zamyslíme s odborníky a s manželi Eklovými z Domova důchodců v Praze 10 nad problémy, které vznikají, když cesta života přichází do období, kdy lidé přestávají být — řečeno s ekonomy — produktivní a dostávají se do situace, kdy potřebují větší či menší péči rodiny, okolí, ale také společnosti.

„Stáří nemá více problémů než dospělost, ale více bijí do očí. Prodloužení života má být odměnou, ne trestem, štěstím a nikoli hrůzou. Svět se pomalu stává světem starých lidí a jeho kulturu lze měřit i tím, jak se o ně stará a jaké stáří jim připravuje, jaké stáří nyní mají," říká v pořadu MUDr. František Hájek. Jak mají žít staří lidé? Jaká práva, jaké touhy mají? Je stáří relativní? Jak se o staré lidi starý socialistické společnosti? Chce starý člověk být sám nebo potřebuje k životu manžela, partnera, rodinu? V pořadu bude řeč ještě o dalších problémech. Manželé Eklovi přijali pozvání do studia na Kavčích horách a budou hovořit o svých zájmech, o tom, jak žijí. Svým vrstevníkům vzkazují: „Stárnoucí člověk je více ohrožen nečinností než cílevědomým pohybem, prospívá mu zájem a radost z výsledků jeho aktivity." dm

(1) 20.00

NEMOCNICE NA KRAJI MĚSTA

Poslední díl seriálu je právě před námi, a tak dnes dejme slovo osobě nejpovolanější, jeho tvůrci, scenáristovi Jaroslavu Dietlovi.

Jak na vás zapůsobil fakt, že budete muset napsat sedm nových dílů Nemocnice?

Moc jsem se těšil, i když jsem nikdy předtím nic podobného nedělal. Chtělo se mi setkat se znovu s postavami, třebaže některé z nich byly už po první části dramaticky dost vyčerpané. Musel jsem vymyslet úplně novou koncepci, nový konflikt, přeskupit postavy, určit, která bude v prvním, která ve druhém plánu.

Poněkud neobvyklá byla mj. skutečnost, že do práce na seriálu zasahovaly dvě dramaturgie — hamburská a pražská — a obě mi sdělovaly názory svých divácких okruhů, které bylo přirozené třeba respektovat. Podle pořadavku hamburské dramaturgické skupiny, na jejíž popud jsem

začal pokračování seriálu psát, dbal jsem zejména na důsledné propojení problémů soukromých s pracovními.

Práci na scénáři mi trochu komplikoval určitý „diktát herců". Nikdy předtím jsem totiž pro konkrétního herce nepsal, jenže tady byly už role předem rozdány, jednotlivé postavy byly už součástí hodnoty seriálu, s tím se nedalo nic dělat. Zkrátka — bylo to nejtěžší práce, jakou jsem kdy dělal. K tomu přirozené přistupoval určitý veřejný popud, měl jsem pocit, že ukázal-li se jednou „český Honza" dobrý pro západoněmecký trh, neměl by ani tentokrát ztratit nic ze svého kreditu.

Jaroslav Dudek nám v rozhovoru řekl, že pokračování seriálu otevřelo šanci postavy charakterově vypreciovat, dostat se jim hlouběji pod kůži. **Kam až jste chtěl v těhle pracích dojít, kam jste chtěl postavy Nemocnice dovést?**

Zobecnil bych to takhle: Potkal jsem před časem, to se ještě ani nezačalo natáčet, na přikopech Josefa Abrháma. Volal na mě — tak co, zase budu zlejt? To se budete divit! volám na něho zpátky. A byl jsem pak svědkem toho, že se skutečně divil — jeho role je právě příkladem možnosti určitého zvratu charakteru předem celkem jednoznačně daného.

Přemýšlíte o nějakém dalším pokračování Nemocnice?

Rozhodně ne. f

Foto Marie Votavová a Otakar Hůrka

(2) 21.45

PROROK, ZLATO A TRANSYLVÁNCI

Dobrodružný žánr, okořeněný trochou humoru a ironizujícím nadhledem autorů, tak bychom mohli lapidárně charakterizovat rumunský film režiséra Dana Pity — Prorok, zlato a Transylvánci. Sám název prozrazuje i hlavní aktéry příběhu; dodejme tedy jen, že se spolu s filmovými hrdiny ocitneme v Americe, v druhé polovině 19. století. Ty, jež lákalo dobrodružství anebo vidina bohatství, směřovali tehdy do míst, odkud se rychle šířila pověst o nalezištích zlata, ke stanici Cedar City ve státě Utah, jehož hlavní město bylo současně střediskem náboženské sekty mormonů. A právě k němu se blíží vlak, v němž se nachází kromě typické westernovské společnosti i podivná dvojice. Prošedivělý Traian s mohutným knírem, oblečený jako boča a jeho mladší bratr Romulus, který neustále, i když s malým úspěchem, hledá v anglicko-rumunském slovníku. Obyčejní rumunští venkované, kteří se vydali za svým bratrem lonem do Ameriky, kam před lety odjel za prací. Komplikace nastávají hned po příjezdu na nádraží. Nejen že se přiletěu k přestřelce mezi „ostré hochy", doplatí i na svou bezprostřednost. Když totiž objeví plakát s fotografií svého bratra, hrdě se

k němu hlásí, neboť nepochopili, že jde o zatykač na lana, známého jako Johny Brad. Rázem se z nich stávají podezřelí a nebezpeční cizinci . .

Divácký ohlas filmu vedl autory k natočení dalších dvou pokračování. Tvůrci nezastírají, že točili především pořad oddechového charakteru a v mnohém se spolehli na osvědčené prvky dobrodružného žánru. Nechávají tu však zaznít i vážnější motivy. Všímají si sociálního pozadí země „neomezených možností" a takzvané zlaté horečky. I když dějová sevřenost filmu poněkud trpí nefunkčním prodlužováním některých akčních scén (například přestřelky), věříme, že sympatičtí hrdinové filmu přispějí k příjemné strávenému večeru u televizní obrazovky. jat

Obrázek B.7: Formát programu pro roky 1979-1990

F1

8.30 — STUDIO ROSA
Sport Billy — Zahrada — Sněhulák
10.00 — Hovory o náboženství
10.15 — Tah sportky
10.25 — SVĚTOVÝ POHÁR VE SJEZDOVÉM LYŽOVÁNÍ
Přímý přenos 1. kola slalomu mužů
(Kitzbühl)
11.15 — Objektiv
Zpravodajsko—publicistický magazín zahraničních zpravodajů
(Premiéra)
11.45 — Událost týdne
11.55 — Minuty dné
12.00 — Co týden dal
Známé osobnosti se zamýšlejí nad uplynulým týdnem
(Premiéra)
12.55 — SVĚTOVÝ POHÁR VE SJEZDOVÉM LYŽOVÁNÍ
Přímý přenos 2. kola slalomu mužů
(Kitzbühl)
13.45 — Vaše hobby — naše hobby
Kalendář, zrcadlo doby — Kouzlo grafických novoročenek — Bylinkami proti nachlazení — Jak vymódit svetr — Čím překvapit hosty — Aby vám za oknem kvetla orchidea. Režie M. Čajko
(Košice)
14.25 — J. Lada
O CHUDÉM KRÁLOVSTVÍČKU
Veselá pohádka o rozpustilé princezně Ance. Hrají: V. Freimanová, Z. Dítě, V. Fialová, S. Zázvorková, P. Štěpánek, M. Stehlík, L. Kostelka a další. Scénář a režie V. Janečková
15.30 — Auto—moto revue
Magazín pro motoristy
16.10 — Nejlepší z Moravy — Šohajka
Hudebně zábavný pořad. Hraje dechový orchestr Šohájka. Připravili M. Duda a M. Peloušek
(Premiéra — připravilo Brno)
16.50 — Minuty dne
17.00 — MS V PLAVÁNÍ
Přímý přenos zpravodajství z průběhu soutěží v Perthu
18.00 — Walt Disney uvádí
Šedesát minut v báječném světě Chipa, Dalea a všech z Kačerova
(Premiéra)
19.00 — Večerníček
Jája a Pája
19.10 — Přehled večerních pořadů
19.15 — Stalo se před rokem
19.20 — Předpověď počasí
19.25 — Reklama
19.30 — Deník ČST
19.50 — HORALKA (2.)
Závěr italského televizního filmu natočeného podle románu A. Moraviі. Scénář D. Gouldová, B. Zapponi, L. Raverová a D. Risi. Režie D. Risi. Hrají: S. Lorenová, S. Pennyová, A. Occhipinti, C. Caloná, L. Ferrantini, D. Ghirardi, T. Givorne a další. Režie českého znění O. Walló
(Premiéra)
21.30 — Reklama
21.35 — O Audienci
Dokumentární film J. Menzela konfrontující známou Havlovu hru s realitou života
(Premiéra)
22.15 — Branky, body, sekundy
22.30 — Joe Cocker
Záznam koncertu
(Premiéra — Bratislava)
23.30 — Zprávy
23.35 — Závěr vysílání

ČTV

10.00 — Křesťanská neděle
Na téma tolerance hovoří duchovní Náboženské společnosti československých unitářů. Připravili M. Duda a Z. Přikaská
(Premiéra — připravilo Brno)
11.00 — KRÁSNÉ ČTVRTI (1.)
Francouzský televizní film podle románu L. Aragona. Scénář P. Savatier. Hrají: J. Guiomar, L. Bellon, J. P. Aumont, B. Brieux a další. Režie J. Krechbron
(Premiéra — v původním znění s titulky)
12.30 — Archa Noemova
Neohrožená zvířata na Galapágách
Španělský přírodopisný seriál
Přestávka od 13.00 do 14.55
14.50 — Entrée libre
10. lekce
15.25 — Entrée libre
40. lekce
15.55 — ALBUM
Čtrnácté vydání zábavného čtrnáctideníku pro pamětníky. Setkání s Josefem Bekem — Čertovská poudačka — Pravidelná soutěž. Připravili A. Ulm, M. Ducháč, A. Doležel, K. Holub a J. Vondráček
(Premiéra)
16.30 — M. Kubátová
ČERTOVSKÁ POUDAČKA ●
Televizní inscenace lidové pohádky. Hrají: J. Bek, J. Bohdalová, B. Záhorský, S. Neumann, O. Nový, V. Sloup a další. Kamera E. Landisch. Režie M. Zachata
18.05 — Setkání s obrazovkou
Týdeník o vysílání ČST
18.25 — Přehled večerních pořadů
18.30 — Dobrý večer, děti
Jája a Pája
18.40 — Reklama
18.45 — Pražský večerník
19.00 — Americká angličtina
30. lekce
19.15 — Pověsti a osudy
19.25 — Reklama
19.30 — Deník ČST
19.50 — Knihy pro vás
19.55 — Přehled večerních pořadů
20.00 — Reklama
20.05 — Koncert Vladimíra Mikulky
Záznam sólového recitálu kytaristy ze Zrcadlové síně pražského Klementina. Kamera I. Bojanovský. Režie I. Kučera
(Premiéra)

Vladimír Mikulka

20.45 — Aktuality
21.05 — Zlatá lyže Vysočiny
Reportáž z průběhu mezinárodního lyžařského závodu
(Premiéra — Brno)
21.35 — DVAKRÁT ZA ŽIVOT
Filmový příběh muže, který chce v padesátce začít nový život. Scénář C. Welland. Režie B. Yorkin. Hrají: G. Hackman, A. Margretová, E. Burstynová a další. Režie českého znění Z. Sirový
23.20 — Závěr vysílání

ST

16.55 — Laura — tulenie mláďa (2.)
Anglický televízny seriál
17.20 — Rhytmick
Prehliadka videoklipov z celého sveta
18.00 — Večerníček
18.10 — Votrelci
Kapre
(Banská Bystrica)
18.40 — Televízny informačný kaleidoskop
19.10 — Hovorí predseda vlády SR
19.30 — Denník ČST
19.50 — Pozývame vás do divadla
J. Palárik: SAMOZVANEC
Prevzaté predstavenie z Divadla P. O. Hviezdoslava v Bratislave
21.50 — Aktuality
22.15 — VRAŽDA PODĽA PREDLOHY
Anglický televízny film
23.10 — Záver vysielania

● — černobílý pořad
⍁ — pro sluchově postižené

NEDĚLE

13. LEDEN

F1 19.50

HORALKA ■ 2

V novém televizním zpracování se změnilo mnohé — kromě tvůrčího štábu v čele s Dinem Risim, hereckého obsazení — J. P. Belmonda v roli Michela vystřídal Andrea Occhipinti, roli dcery Rosetty svěřil režisér mladé Sidney Pennyové, se změnil i přístup k Moraviovu dílu. Čtveřice scenáristů D. Gouldová, B. Zapponi, L. Raverová a D. Risi se na rozdíl od Zavattiního a De Sikova pojetí držela daleko více literární předlohy, a tak se v televizní verzi objevily i některé další motivy, které v původním filmovém zpracování chyběly.

Co se však po osmadvaceti letech nezměnilo, je právě představitelka hlavní role, kterou Dino Risi opětovně svěřil stále půvabné, dnes již však o mnoho zkušenější a vyzrálejší Sophii Lorenové. »Když mi nabídli, abych znovu hrála v Horalce, nesmírně mě to překvapilo,« říká Sophie Lorenová. »Nikdy jsem na to ani nepomyslela. Když však přišli znovu, mé překvapení narůstalo. Nepředpokládala jsem takové naléhání. Začala jsem o tom tedy přemýšlet a ptala jsem se sama sebe proč. Po týdnu jsem dospěla k závěru, že je to vlastně výborný nápad. Mohu si porovnat svůj život tehdy a dnes, jaká jsem byla a jaká jsem, je to taková výzva mému mládí. Tehdy byly mé vzpomínky na válku ještě velmi čerstvé, dnes již trochu vyprchaly. I já sama jsem se výrazně změnila, stala se ze mne konečně matka ... Uvěřila jsem tedy, že se mohu opět zhostit svého úkolu dobře a obohatit postavu Cesiry i o všechny své zkušenosti, které jsem za ta uplynulá léta nashromáždila . . .«

Českou verzi pořadu s vlastními dialogy a s použitím překladu J. Hampacherové natočila režisérka O. Walló. Hlavním postavám své hlasy propůjčili R. Merunková, D. Prachař, E. Spoustová a další. ap

ČTV 11.00

KRÁSNÉ ČTVRTI

S dvouletým zpožděním přichází na naše obrazovky třídílný francouzský seriál Krásné čtvrti, natočený podle románové předlohy Louise Aragona. Pro ČST byl seriál s jistým rizikem přijímán v době, kdy Aragon byl u nás ještě tabu, kdy jeho spisy byly nežádoucí stejně jako zmínky o jeho osobě. Bylo to absurdní, ale co v dobách normalizace absurdní nebylo? Vždyť Aragon byl jednoznačně levicově orientovaný autor, od roku 1927 člen Francouzské komunistické strany a od roku 1956 dokonce člen jejího ÚV. Stejně jako většina západoevropských komunistických stran se i on, věren svému důslednému antifašistickému postoji a neúnavné veřejné aktivitě, vyjádřil negativně k okupaci Československa v roce 1968, a to byl kámen úrazu . . . Louis Aragon se narodil v roce 1897 a zemřel v roce 1984. Zanechal po sobě neobyčejně rozsáhlé dílo: sbírky básní, romány, povídky, eseje, kritické statí, překlady z ruštiny. Pro starší generaci zůstávají mimořádné vzrušivá léta vrcholné surrealistické tvorby, objevů a experimentace tohoto dráždivého hnutí, k němuž se ve dvacátých letech Aragon od dada spontánně přiklonil. Není bez zajímavosti, že i před touto částí básnikovy tvorby byli svého času naši čtenáři jako před čímsi dekadentním, škodlivým administrativně »chráněním«. Nebylo žádoucí poznání, že to byl právě surrealismus, který více než kterékoliv jiné umělecké hnutí našeho století prokázal, že má své oprávněné gnoseologické předpoklady a že právě jemu šlo především o postižení lidské stránky skutečna. Odtud i logika Aragonovy cesty k monumentální činnosti prozaistické, protože, málo platné, pro básníka je skutečnosti přítomnost. A svět se od časů mladického Aragona příliš změnil a prošel příliš složitými, podivuhodné klikatými obdobími. vk

ČTV 21.35

DVAKRÁT ZA ŽIVOT

Padesát let zaběhnutého života, pečlivá manželka, hodné dospělé děti, pravidelné a fyzicky náročné zaměstnání — to vše by člověka mohlo vést k uspokojení a klidu. Ale co když se muž ještě cítí plný síly a chce se chopit náhodné příležitosti, která se mu naskytne, aby dokázal, že ještě nepatří do starého železa? Pak samozřejmě udělá rozhodnutí, jímž způsobí mnoho bolesti všem lidem, které měl dosud rád. Situace některak originální, dokonce by se dalo říci, že v mnoha případech běžná, je základním schématem filmu USA nazvaného příznačně Dvakrát za život, který v roce 1985 natočil režisér Bud Yorkin. Scenárista Colin Welland se v něm zamýšlí nad složitými rodinnými vztahy, podobně jako to činily jiné filmy, např. Obyčejní lidé, Cena za něžnost a podobně. A stejně jako v nich si pozornost diváka zaslouží zejména herecké výkony, vždyť hlavní roli dělníka Mackenzieho vytvořil nám dobře známý Gene Hackman (Francouzská spojka, Domino, Rozhovor, Pod palbou). Jeho novou partnerku hraje zpěvačka švédského původu Ann Margretová, u nás známá např. z filmu Tygr a kočička (Itálie) či Dostavník.

Film uvede v reprize na obrazovku Hlavní redakce pořadů ze zahraničí s českou verzí dabingového studia Čs. státního filmu. Hlavním postavám propůjčili hlasy I. Racek, M. Drahokoupilová, V. Galatíková, J. Drbohlavová a G. Osvaldová. Dialogy a režie jsou dílem Z. Sirového. Televizní premiéra se konala v květnu loňského roku. r

Obrázek B.8: Formát programu pro roky 1990-1992

Příloha C

Ukázka televizního programu z týdeníku Rozhlas a televize

Obrázek C.1: Ukázka nedostatečně kvalitního televizního programu z týdeníku Rozhlas a televize z roku 1962, který nebyl naskenován, ale pouze vyfotografován a text je proto rozmazaný