

PILLER IMRE¹¹

AZ ANYAKÖNYVEKBEN SZEREPLŐ ADATOK HATÉKONY KEZELÉSE GÉPI LÁTÁS
SEGÍTSÉGÉVEL

Absztrakt

Az elektronikus anyakönyvi rendszer használata ellenére továbbra is vannak olyan adatok, amelyek az anyakönyvvezetők számára csak az eredeti anyakönyvekben, papír alapon érhetők el. A dolgozat célja, hogy bemutassa azt, hogy a gépi látásra épülő módszerek segítségével hogyan oldható meg az összes szükséges adat elektronikus formában történő kezelése. A dokumentumok digitalizálásán túl foglalkozni kell az aktuálisan használt rendszerekkel való integrációval. Fontos azt megnézni, hogy az anyakönyvvezetők mely feladatainál és hogyan tudnak segíteni a gépi látás és gépi tanulás eszközei. A dolgozat áttekinti az ezek megvalósításához szükséges feltételeket, továbbá javaslatokat tesz a kivitelezés lépéseire vonatkozóan.

Kulcsszavak: anyakönyv, gépi látás, gépi tanulás, információkezelés

EFFICIENT MANAGEMENT OF VITAL RECORDS USING COMPUTER VISION

Abstract

Despite the use of the electronic registry system, there is still data available to registrars only in the original registers on paper. The goal of this work is to present how computer vision-based methods can be used to manage all required data in electronic form. Beside the general problems of digitalization, we have to consider the possible ways of the integration with the existing systems. It is important to mention, how can the modern tools of computer vision and

¹¹ Miskolci Egyetem, Matematikai Intézet, Alkalmazott Matematikai Intézeti Tanszék, egyetemi tanársegéd

Az anyakönyvekben szereplő adatok hatékony kezelése gépi látás segítségével

machine learning help the daily work of the registrars. The paper reviews the requirements and the proposed steps of its implementation.

Keywords: vital records, computer vision, machine learning, information management

1. Bevezetés

Az anyakönyvvezetésnek hagyományosan elengedhetetlen eleme volt maga az anyakönyv. Ezek szerepükben, kivitelezésükben és fizikális terjedelmükben is tekintélyt parancsoló dokumentumok. Az anyakönyvvezetők mindennapjainak manapság is szerves részét képezik, bár jelentőségük az Elektronikus Anyakönyv (röviden EAK) bevezetésével csökkent. Egyfajta átállási folyamatnak lehetünk tanúi, melyben az új technológiák fokozatosan kiszorítani látszanak a korábbi évtizedek gyakorlatait.

A dolgozat azt mutatja be, hogy ezen digitalizációs folyamatban milyen szerep juthat a korszerű, mesterséges intelligenciára és gépi tanulásra épülő módszertanoknak. Igyekszik egyaránt figyelembe venni az új tudományos eredményeket, a rendelkezésre álló technikai háttérrel, illetve az anyakönyvvezetés aktuális módját is, mivel utóbbi elengedhetetlen ahhoz, hogy realitása legyen a felvázolt módszerek alkalmazhatóságának.

A dolgozat négy fő tartalmi részből áll. Először a digitalizáció módját és jelentőségét mutatja be, mivel az anyakönyvek számítógéppel való hozzáféréseinek a biztosítása szükséges előfeltétele a továbbiaknak. Az ezt követő fejezetben magáról a felvetett átállási folyamatról olvashatunk, azt szintekre, fázisokra bontva. A gépi tanulás, mint a mesterséges intelligencia módszereinek egyike, a dokumentumok automatizált feldolgozásában segít. Egy külön fejezet foglalkozik azzal, hogy a képként rendelkezésre álló anyakönyvi bejegyzésekből milyen módszerekkel lehet majd az elektronikus rendszerek számára kezelhető formátum. Végül a kivitelezés feltételeinek, tervezett lépéseinek áttekintését láthatjuk.

A dolgozatnak nem célja, hogy részletezze az anyakönyvvezetés múltját és jelenét. Az anyakönyvekben szereplő adatok hozzáférhetővé és elemezhetővé tételével foglalkozik. A mesterséges intelligencia alkalmazásának technikai jellegű részletezésébe csak olyan szinten megy bele, amely annak belátásához szükséges, hogy a tervezett módszer valóban előnyös és használható.

Az anyakönyvben szereplő adatokhoz való hozzáférés jogosultságainak törvényi rendszerének taglalása szintén nem része a dolgozatnak, ahogyan az EAK-kal való integráció pontos módjának leírása sem. Egy olyan koncepcióról van szó, amely megvalósítása esetén jelentősen egyszerűsíteni és gyorsítani tudja az anyakönyvvezetők munkáját.

2. A digitalizálás módja és jelentősége

Az adatok digitalizálása és elektronikus kezelése egyértelműen láthatóan fokozatosan felváltja a papír alapú adatkezelési módszerek használatát. A digitalizálás különféle célból, változatos eszközkészlet felhasználásával valósulhat meg. Ennek módja, nehézségei, esetleges problémái aktuális kutatások tárgyát képezi nemzetközi szinten. Egy olyan átállásról van szó az anyagok digitális archiválása kapcsán, amely technológiai háttére adott, az akadályt elsődlegesen már az emberi tényező jelenti (*Perry 2014*).

A közigazgatásba az új, elektronikus technológiák bevezetése évtizedek óta zajlik, mivel rengeteg sajátos, más területeken nem jelentkező problémát vet fel (*Yildiz 2007*). Az anyakönyvben szereplő adatok kezelése, az azok digitalizálásával, és automatikus feldolgozásával kapcsolatos vizsgálatok nyilvánvalóan csak egy szeletét jelentik az e-közigazgatás rendszerének. A dolgozat ezt, mint egy konkrét esetet vizsgál, de a felvetett gépi tanulásra épülő felhasználási módok minden olyan kapcsolódó területen jól alkalmazhatóak, ahol az adatok jelenleg papír alapon érhetők el valamilyen jól meghatározott formában (például űrlapokon).

A következőkben annak a vizsgálatára kerül sor, hogy az anyakönyvi adatok esetében melyek lehetnek ezek a célok, és milyen módokon, módszerekkel valósulhatnak meg.

2.1. A digitalizálás lehetséges céljai

Az anyakönyvek alapvetően kézzel írott, papír alapú dokumentumok. A benne lévő részek egymástól (elvileg) jól elkülöníthető módon, táblázatos formában, vonalakkal elválasztva találhatók meg.

Az anyakönyvekben szereplő adatok hatékony kezelése gépi látás segítségével

The image shows two pages of a handwritten church register. The left page is titled "Házasságok anyakönyve" (Marriage Register) and the right page is titled "a [paróki] i egyházban, 1844 évről" (in the [parish] church, 1844). Both pages contain tables with columns for names, dates, and other details, filled with handwritten entries.

1. ábra: Anyakönyvi bejegyzés

Forrás: Magyar Nemzeti Levéltár, weboldal: <http://mnl.gov.hu>

Anyakönyvi bejegyzésekre láthatunk példákat az 1. ábrán. Ez ugyan nem egyezik meg sem tartalmában, sem pedig formátumában azzal, mint amelyeket az anyakönyvvezetők használnak. A példa olyan szempontból viszont helytálló, hogy az oldalak sajátosságai, a táblázatos elrendezés hasonló, és a példában szereplő adatok használatával kapcsolatban sem merülhetnek fel jogi problémák, mivel azok publikusan elérhetők.

A digitalizálásuknak általánosan a bennük lévő adatok hozzáféréseinek a segítése és az írásos dokumentum hosszabb távú megőrzése, archiválása lehet a célja. A következő alpontokban néhány konkrétabb szempontot vizsgálunk meg, melyek már felvetés szintjén utalnak majd a mesterséges intelligencia lehetséges alkalmazási módjaira.

2.1.1. Fizikai hozzáférés nélkülözhetővé tétele

Az anyakönyvekkel való munkát kézenfekvő módon nehezíti az, hogy a könyvek nagyok és nehezek. Általában egyidejűleg több ügy intézése is folyamatban van, amelyekhez tartozó bejegyzések külön anyakönyvekben szerepelnek. Forgalmasabb időszakban a nagyobb anyakönyvi hivatalokban hatalmas tornyokba rendezve láthatjuk ezen dokumentumokat. Az aktuálisan hozzáférhető technológiai szint mellett már lehetőség lenne ezen dokumentumok elektronikus formában való elérésére. Az elektronikus anyakönyvi rendszerek nyilvánvalóan

többek között ezt a problémát igyekeznek megoldani. Teljesen kiváltani viszont azért nem tudják még a papír alapú anyakönyveket, mert a bennük lévő bejegyzések csak egy része került még rögzítésre. Elképzelhető olyan megoldás (mesterséges intelligencia nélkül is), amely számítógépen keresztül teszi elérhetővé a dokumentumokat. Ekkor az anyakönyvvezető a könyvekben és a bejegyzések között való keresést a szokásos módon végezheti, viszont így már nem szükséges magának a könyvnek helyben elérhetőnek lennie.

2.1.2. Hatékonyabb keresés

A bejegyzések kereséséhez sajnos gyakran nem áll rendelkezésre minden szükséges információ. Amennyiben például egy dátum nem ismert, akkor bizonyos intervallumon belül kell kézzel végignézni a bejegyzéseket. A dátumok aránylag kötött, rendezett formában található meg az anyakönyvben, így azok felismerése nagy megbízhatósággal elvégezhető. Adott időintervallumhoz tartozó bejegyzések úgy is lekérdezhetővé válnának, hogy azok külön anyakönyvekben található meg.

Gépi látás segítségével becslés adható a bejegyzésekben lévő adatok egy részére. Például, hogy ha ismert egy névtöredék, amely alapján a bejegyzést meg kell keresni, akkor ahhoz már elérhetőek olyan optikai karakterfelismerő alkalmazások (*OCR - Optical Character Recognition*), amelyekkel keresni/szűrni lehet a szóbajóhető bejegyzések között.

Ahhoz, hogy az a kézzel írott szövegben géppel szöveg alapján keresni lehessen, azt először értelmeztetni kell. Írott szövegek esetében a Google-nek is van egy szabványa, amely segítségével ez elvégezhető (*Al-Karmi et al. 1999*). Természetesen számos más módszer is rendelkezésre áll, magán a mesterséges intelligencia és a gépi tanulás témakörein belül is. A dolgozat ezen módszerekkel csak ötlet és hivatkozás szintjén foglalkozik.

2.1.3. Automatizált adatfelvitel

Távlati célként tekinthetjük azt, hogy minden anyakönyvi bejegyzés számítógépen keresztül elérhető legyen. A gépi látás és a gépi tanulás alkalmazásával konstruálhatók olyan algoritmusok, amelyek a kézzel írott szöveget automatizált módon feldolgozzák, és az eredményeket strukturált formában adják vissza. Ez már mindenképpen egy komplikált, statisztikai jellegű becslésekre támaszkodó művelet. Az ilyen rendszerek nem tudnak hiba nélkül működni, így azt várhatjuk el, hogy a gép a képek formájában tárolt anyakönyvekből

képes kinyerni a bennük lévő adatok egy részét, majd azt az anyakönyvvezetők által jóváhagyva vagy korrigálva át tudja emelni az elektronikus rendszerbe.

2.1.4. Írásos dokumentum megőrzése

Az anyakönyvek olyan írásos dokumentumok, amelyeket mindenképpen meg kell őrizni. A felvetett digitalizálási és elektronikus rendszerekre való átállás ugyan a hétköznapi ügyintézés szempontjából nélkülözhetővé tenné őket, viszont a történelmi hagyaték ápolásaként mindenképpen vigyázni kell rájuk. A képi formában tárolt bejegyzéseket az után is érdemes megőrizni, hogy a bennük lévő adatok már átemelésre kerültek. A digitalizálás tehát nem ezen dokumentumok ellen, hanem megőrzésük érdekében is használható.

2.2. A digitalizálás módjai

A dokumentumok fényképezése manapság már egyáltalán nem jelent problémát. A gépi látás esetében érdemes bizonyos szempontokat szem előtt tartani (mint például a kép megfelelő felbontása, a fényviszonyok), de már a kereskedelmi forgalomban kapható fényképezők és lapolvasók is alkalmasak ilyen feladatokra.

A digitalizáláshoz egyszerűen összeállítható megfelelő eszköz. Mindössze egy fényképezőre, egy állványra és a megfelelő fényviszonyok biztosításához egy lámpára van szükség.

Léteznek olyan, nagyobb méretű fekvő ágyas lapolvasók is, amelyeknek nem jelent problémát egy anyakönyv méretű dokumentum szkennelése. Ezek használata azért nem tűnik praktikus döntésnek, mert jóval költségesebbek, mint az előző bekezdésben felsorolt eszközök, képminőségben nem jelentenek számottevő előnyt, és a szkennelésnél az anyakönyvek lapozását is körülményesebbé teszik.

2.3. Lehetséges problémák

Nyilvánvalóan a digitalizált és gépi tanulással támogatott rendszer használatának is meg lehetnek a hátrányai. Ezek részletes áttekintése külön hatástanulmányt igényelne. A teljesség igénye nélkül a következőkben két lehetséges problémára hívnánk fel a figyelmet.

Mint minden új rendszernek, ennek az elfogadásával is problémák adódhatnak az anyakönyvvezetők részéről. A számítógépen való elérésre való átállást nem feltétlenül fogadják örömmel minden esetben (főként az idősebb korosztály). A kézzel fogható anyakönyveknek van egy olyan presztízse a hivatalokban, amelyet semmilyen rendszer nem fog tudni pótolni. A bizalmatlanságra további okot adhat, hogy időbe telik, amíg egy rendszer kiforrottá válik, és a megoldott régi, napi szintű problémák mellett újakat is felvet.

A digitalizált rendszer akkor lehet hatékony, hogy ha megfelelő módon integrálva van az aktuálisan használt elektronikus anyakönyvi rendszerrel. Ez technológia szintjén nyilván megoldható, de külön ráfordítást igényel az informatikusok részéről.

3. Integráció az aktuális rendszerekkel

A rendszerekbe itt beleérthetjük a napi gyakorlatban használt módszereket is. A fejezet lehetséges forgatókönyvekkel illusztrálva tekinti át azt, hogy a különböző megvalósítási szintek hogyan illeszkedhetnek ezekhez. Tulajdonképpen az átállást egy többszintű folyamatként képzelhetjük el, amely már megkezdődött. Itt azt vizsgáljuk meg, hogy a modern eszközök (értve ez alatt egyaránt hardvert és szoftvert is) hogyan segítheti ezt.

3.1. Elektronikus anyakönyvi rendszer használata

Az anyakönyvezés modernizálására való törekvések egyáltalán nem újkeletűek. Már a 90-es évek közepén előrevetítették, hogy melyek azok a területek és használati módok ezen belül, amelyek segíthetik a hatékonyabb adatkezelést. Az elektronikus születési kivonat (*Electronic Birth Certificate*) bevezetése egyaránt előnyös volt a hivatalok és az egészségügyi nyilvántartások számára is az Amerikában végzett vizsgálatok alapján (*Starr–Starr 1995*).

Úgy tekinthetjük, hogy jelenleg az Anyakönyvvezetői Hivatalok az éppen taglalt fázisban vannak. Az Elektronikus Anyakönyv (röviden EAK) az átállásnak egy olyan változatát valósítja meg, amelyben az aktuális ügyek intézéshez szükséges adatok kerülnek rögzítésre. A hosszabb távú cél ez alapján semmiképpen nem lehet az, hogy biztosan minden

elérhető legyen digitalizált formában. Az könnyen látható, hogy az anyakönyvek jelentősége így az idő múlásával fokozatosan ki tud szorulni, viszont ez egy hosszabb időintervallumot ölel fel. A folyamat végét az jelzi majd, hogy ha a papír alapú dokumentumokból való keresésre már csak elvétve, vagy ideális esetben egyáltalán nem lesz szükség. Önmagában ez a stratégia tehát nem teszi lehetővé, hogy a következő évtizedekben ne legyen szükség könyv formájában az anyakönyvi bejegyzésekre.

3.2. Digitalizált repozitóriumban való "kézi" keresés

Tegyük fel, hogy minden anyakönyv digitalizálásra került, és a bennük lévő hely és idő adatok alapján rendszerezett formában elérhetők. Az anyakönyvvezető munkája így olyan formában egyszerűsödik, hogy a könyvben való lapozás helyett azt a számítógépen keresztül megteheti.

A korábbiakban nem került hangsúlyozásra, de már ez a változat is számos járulékos előnnyel rendelkezik. Az egyik lényeges az, hogy a könyvek így már nem lesznek helyhez kötve. Egy központi, online elérhető rendszert feltételezve (amelyhez hasonlókat nyilvánvalóan az ügyek intézéséhez egyébként is használnak), a települések anyakönyvei távolról is elérhetővé válnak. Az adatok egyeztetésénél gyakran pont ez a hozzáférési probléma jelenti a szűk keresztmetszetet.

További előny, hogy a kérdéses bejegyzésekkel, hibás adatokkal kapcsolatban egyszerűbb egyeztetni. Amennyiben például egy bejegyzésből nem derül ki egyértelműen, hogy a névben milyen betűk szerepelnek, akkor azt az akár ilyen jellegű problémákra szakosodott kollégákkal a rendszeren keresztül könnyen meg lehet tenni.

Az anyakönyvvezetők szemszögéből itt kifejezetten egyszerű felületekre gondolhatunk. Tulajdonképpen ez a szint csak a fizikális könyvek kiváltását hivatott megoldani.

3.3. Összetett keresés alkalmazása

A gépi tanulós módszerek bizonyos megbízhatósági szinten tudják felismerni a kézirásos szövegeket. Elképzelhetők olyan kereső algoritmusok, amelyek a felismerés megbízhatóságának megfelelően adják vissza az eredményeket. Definiálhatók lennének így olyan összetett keresési feltételek, amelyekkel úgy is meg lehet találni egy-egy bejegyzést, hogy ha az a feltételekre pontosan nem illeszkedik.

Az anyakönyvvezetőknek egy ilyen jellegű rendszer használata már külön előkészületeket, felkészülést igényelne. Nehéz azt megbecsülni, hogy a részinformációk ismeretében hogyan érdemes egy lekérdezést megfogalmazni, mikor jobb szigorúbb szűrési feltételeket megadni, és mikor előnyösebb, hogy ha a kiadott több találatot kell egyesével végignézni.

3.4. Gépi tanulással segített adatfelvitel

Az írott szövegek felismerése a gépi látás egy speciális területe. A kézírás önmagában egy komplikáltabb feladat, viszont a folyóírás, az anyakönyvek esetében gyakorta látható cikornyás folyóírás ezt még bonyolultabbá teszi. Hivatalos adatokról lévén szó, azt nem várhatjuk el, hogy a gép ezt a problémát olyan megbízhatósági szinten fogja tudni elvégezni, hogy a kapott eredmények közvetlenül rögzíthetők. Mindenképpen egy olyan rendszerben szabad itt gondolkodni, amelyik az automatikusan feldolgozott képből kinyert adatokat az anyakönyvvezetővel hagyatja jóvá.

A hivatali gyakorlatban ezt úgy képzelhetjük el, hogy az anyakönyvvezető bizonyos szempontok alapján keres egy bejegyzést, majd a rendszer erre többféle választ adhat.

- Amennyiben az adatok egyértelműen leírták a találatot, és az már korábban jóvá lett hagyva (vagyis egy anyakönyvvezető már ellenőrizte, vagy eleve az elektronikus rendszerbe újonnan vitte már fel), akkor további különösebb teendője nincs vele.
- Amennyiben az adott feltételekhez csak olyan találat van, amelyekhez még nincs ellenőrzött bejegyzés, akkor az anyakönyvvezetőnek a kép formájában elérhető bejegyzéseket kell végig néznie, amelyekhez viszont már szöveges, strukturált formában is rendelkezésre állnak a feldolgozott adatok. A feladat ekkor az, hogy összehasonlítsa, jóvá hagyja, szükség esetén pedig korrigálja a gépi intelligencia által adott eredményeket.
- Előfordulhat olyan eset is, amelyben jóváhagyott és nem jóváhagyott bejegyzések vegyesen fordulnak elő. Ekkor nyilván az anyakönyvvezetőnek különböző feldolgozottsági szintű adatokból kell kiválasztania az éppen szükséges elemet.
- Esetként tekinthető az is, hogy ha nincs találat a megadott feltételekre, de akkor nyilván a keresési feltételeken kell valamilyen formában lazítani.

Az anyakönyvvezetők szempontjából ez a szint már újszerű feladatokat vet fel. A gépi látásra épülő mesterséges intelligencia igyekszik megkönnyíteni a munkáját azzal, hogy kikövetkezteti a képen szereplő adatok egy részét. Az anyakönyvvezető előtt tehát egy olyan

felület lenne látható, amelyik szegmentált/annotált formában mutatja, hogy a kép melyik részére milyen becslés történt. Ebbe bele kell tudni szerkeszteni, mivel itt feltételezhető, hogy szükség lehet a korrekciókra. Ezt követően a kinyert és ellenőrzött adatok már átemelhetők a szokásos, elektronikus anyakönyvi rendszerbe.

3.5. EAK integráció

Az aktuálisan használt EAK-ba az adatok átemelése külön előkészületeket igényel informatikai részről. Az anyakönyvvezetők munkáját úgy tudja a rendszer hatékonyan segíteni, hogy ha az ellenőrzött adatok egy kitöltött űrlap formájában jelennek meg, amelyekhez esetlegesen plusz információk rendelkezhetők (amelyek magában az anyakönyvi bejegyzésben nem voltak benne, de valami miatt szükség van rájuk). Ez az integráció az elterjedt adattárolási formátumok és átviteli protokollok mellett nem szabad, hogy problémát jelentsen.

3.6. A teljesen elektronikus rendszer

Egy idealizált állapotnak tekinthetjük azt, hogy ha az anyakönyvvezetés rendszere teljesen elektronikussá válik. Azt kellene hozzá elérni, hogy minden anyakönyvi bejegyzés már le legyen ellenőrizve. Ekkor az említett, speciálisan az anyakönyvvezetők számára készített ellenőrző felület egyúttal el is veszti a funkcióját. A bejegyzésekben való keresés, integritásuk ellenőrzése teljesen automatikussá válhat.

A gépi látásra épülő mesterséges intelligencia módszerekre szintén nem lenne már ezen a ponton szükség. Ezen észrevétellel is hangsúlyozandó tehát, hogy a felvetett digitalizálási és feldolgozási folyamat az átállást igyekszik segíteni. Természetesen a mesterséges intelligencia az anyakönyvvezetés más területein is tud még segítséget jelenteni, de az már nem ezen dolgozat témája.

4. A gépi tanulás alkalmazásának módjai

A gépi látást és hozzá kapcsolódóan a gépi tanulást hasonló problémák megoldására jónéhány évtizede már sikerrel alkalmazzák. Illusztrációképpen csak néhány területet említünk ezek közül.

- A postai és banki rendszerekben nagy mennyiségű űrlap feldolgozására van szükség. Mivel ezek formátuma eléggé kötött, így gépi látás segítségével elkerülhető az, hogy minden esetben az adatokat embereknek kelljen ellenőrizniük. Jellemzően az ilyen jellegű űrlapokat nyomtatott betűkkel kell kitölteni, amely segíti, hogy a gépi feldolgozás megbízhatóbb eredményt tudjon adni. A képfeldolgozási módszerek aktuális fejlettségi szintje miatt ezt már kevésbé kutatják. Az OpenCV (*Open Computer Vision*) nevű függvénykönyvtár széles palettáját kínálja a kötöttebb űrlapok feldolgozásához szükséges eszközöknek (*Bradski–Kaehler 2008*).
- Rég óta foglalkoznak, és most már tankönyvi példaként említik a rendszámok/hatósági jelzések felismerését. A közúti ellenőrző rendszerek esetében olyan szoftvereket használnak, amelyek a fényképeken képesek megkeresni a járműveket és a rajtuk lévő rendszámot. Az eszközök és módszerek jól ismerteknek tekinthetők, kutatások, konkrét számítógépes alkalmazások és szabványok formájában is elérhetők (*Silva–Higgins 2019*).
- Napjainkban már szkennelt dokumentumokban tartalom alapján is lehet keresni. Ennek használhatósága erősen függ a kép minőségétől, a nyelvtől és nyilván az alkalmazott algoritmusoktól. Egy ide vonatkozó kutatás pontosan a hasonló, kézzel írott archív anyagok feldolgozását célozza (*Schomaker 2016*).
- A kézzel írott szövegek gépi látással való értelmezése szintén aktívan kutatott terület. Ehhez kapcsolódóan külön vizsgálják a képből kinyerhető grafológiai információkat. A megoldáshoz az egyik szóba jöhető algoritmus az alapvonalak detektálására épül (*Bal–Saha 2016*).
- A dokumentumokban lévő táblázatok felismerésével aktuálisan mély tanulós (*Deep Learning*) vizsgálatokat is végeztek (*Schreiber et al. 2017*).

A következőkben azt fogjuk megvizsgálni, hogy a hasonló jellegű kutatások és alkalmazások eredményeit hogyan, milyen módszertan szerint tűnik érdemesnek felhasználni a tervezett rendszer megvalósításához.

4.1. A probléma modellje

A gépi tanulással megoldandó probléma röviden úgy foglalható össze, hogy a szétnyitott anyakönyvről készített fényképből szeretnénk strukturált adatokat kinyerni szöveges formában.

Hatalmas előnyt jelent az, hogy az anyakönyvek jól meghatározott sémát követnek. Az egyes anyakönyvi kiadások között természetesen vannak eltérések, de a feldolgozási folyamat komplexitására azok nincsenek hatással. A kötött séma az elvárt eredmény szempontjából is előnyös. Előre lehet tudni ugyanis, hogy melyek azok az adatok, amelyekre szükségünk van. Többségében ezek szöveges értékűek. Előfordulhatnak benne számok, dátumok és szimbólumként kezelendő aláírások, de mindegyik esetében lehet tudni, hogy milyen típusú adatra hol számíthatunk.

A feldolgozáshoz a színek hasznos információt szolgáltathatnak. Így feltételezhetjük azt, hogy egy rögzített felbontású színes kép lesz a bemenet. A kimenet az adott sémának megfelelő rekord lesz.

A probléma megoldását egy olyan algoritmus adja, amely képes a bemeneti kép alapján meghatározni a bejegyzésben szereplő adatokat. A modellt annyival még érdemes pontosítani, hogy az algoritmus az aktuális bemeneti képen kívül a könyv formátumára vonatkozó explicit információkat és más anyakönyvi felvételeket is felhasználhat, amelyeknél az is ismert, hogy mely könyv melyik részéből származnak.

4.2. A feldolgozás lépései

Nem szükségszerű, de célszerű a felvételek készítésénél arra törekedni, hogy a képek felbontása, a fényviszonyok és a képen az anyakönyv nagysága minden felvétel esetén a lehetőségekhez mérten közel azonos legyen. A feldolgozás lépéseit attól a ponttól kezdve vizsgáljuk, hogy a kép már digitális formában elérhető.

- A képfeldolgozásnál jellemzően egy előfeldolgozási fázissal kezdődik a képen szereplő információk kinyerésének a folyamata. Ez több, különböző előkészítési módot foglal magába. Először az intenzitások kiegyenlítésére, intenzitáshisztogramok segítségével végrehajtható normalizálási műveletekre van szükség. Ez azokat a hibákat korrigálja, amelyek az eltérő megvilágításból és a lapok színéből, háttérmintázatából adódhatnak. Ezt követően arra van szükség, hogy a kép elemzésre szánt része ki legyen jelölve. Mivel a digitalizálás egy felügyelt folyamat, így egyszerűen elvégezhető, de fontos lépés az, hogy a kép lényegi információkat tartalmazó része ki legyen jelölve. Ez jelentheti a kép széleinek a levágását, vagy pedig a lapok szélének a detektálását.
- A feldolgozandó kép egy táblázat, amely a könyv közepénél két részre van bontva. Képfeldolgozási algoritmus segítségével meg kell tudni határozni, hogy pontosan hol

van az elválasztó vonal, illetve már ekkor hasznos lehet a későbbiekre nézve a fejléc, oszlopfeliratok és az oldalszám megkeresése is.

- A bejegyzések a táblázat soraiban találhatóak. Ezek száma, az egyes cellák jelentése jól meghatározott. Azokat érdemes magába a modellbe beépíteni. A feldolgozási lépés eredményeképpen a bejegyzések sorait el kell választania az algoritmusnak egymástól.
- A bejegyzések szerkezete szintén jól definiált. A következő lépésben meg kell keresni a bejegyzésen belül, hogy melyik szövegrész hol található.
- A képen kijelölt szöveges információkat tartalmazó részekre külön-külön elvégezhető az optikai karakterfelismerés (vagy ahhoz hasonló írott szövegre optimalizált szövegfelismerési módszer). Ez már karaktereket, szótöredékeket, szavakat, sorokat eredményez, amelyekhez még hozzátartoznak a képen lévő pozíciók.
- A feldolgozás addigi törmelékes eredményeiből összevonások és következtetések segítségével már össze lehet állítani a bejegyzés szöveges tartalmát.
- A szöveget érdemes természetes nyelvi feldolgozási módszerekkel validálni, tisztítani, kiegészíteni.

Nyilvánvalóan érezhető, hogy a feldolgozási lépések így logikusan következnek egymás után, de számos, még önmagában is nehezen kezelhető részproblémát rejtnek magukban. A következő szakaszokban a várhatóan előforduló gépi tanulási problémákról, és azok lehetséges megoldási módjairól olvashatunk.

4.3. Bejegyzések elkülönítése

A bejegyzések elkülönítése a megoldandó problémának egyszerűbb része. Ehhez az anyakönyv vonalazása és a rögzített helyen lévő információk nagy segítséget jelentenek. A következő alpontok azt taglalják, hogy ezeket hogyan tudjuk hatékonyan felhasználni.

4.3.1. Határoló vonalak detektálása

Az anyakönyvek formátuma szándékosan olyan, hogy a bejegyzések egyes részei jól látható módon, vonalakkal vagy keretekkel el legyenek választva egymástól. A digitalizálásnál ideális esetben ezek függőleges és vízszintes vonalak lesznek, de kisebb elforgatást is könnyen korrigálni lehet. A határoló vonalak detektálását akár az adott anyakönyv formátumának megfelelő sablon illesztésével, vagy vonalkereső algoritmusok alkalmazásával is nagy pontossággal meg lehet találni.

4.3.2. Rákövetkezési információk

Az oldalakon található oldalszámok és dátumok mind utalnak arra, hogy a bejegyzések milyen sorrendben következhetnek egymás után. A bejegyzések elkülönítésénél, esetlegesen törölt bejegyzések felismerésénél ezek is fontos információkat hordoznak. A rákövetkezés egy-egy teljes anyakönyv adatainak integritásellenőrzéséhez is használhatók.

4.4. Bejegyzések adatainak elemzése

A probléma komplikált részét az jelenti, hogy hogyan tudjuk a szöveges részeket tartalmazó régiókat kijelölni a képen. Ez az a pont, ahol már mindenképpen indokolt valamilyen mesterséges intelligencia, vagy konkrétan gépi tanulási módszert alkalmazni. (Korábbi, a gépi tanulásos módszerek széles körben való elterjedése előtt is már foglalkoztak kézzel írott szövegek elemzésével (*Rodrigues–Thomé 2000*). Az eredményeik alkalmazhatóak ugyan, de várhatóan a gépi tanulás nagyobb mintaszám esetén pontosabb eredményt fog tudni adni.)

Tipikus megközelítés, hogy olyan algoritmust készítünk, amely bizonyos képrészletről képest azt megbecsülni, hogy azon található-e szöveg vagy nem. Konvolúciós mesterséges neurális hálók alkalmazásával minden képpontról megállapítható, hogy az szöveghez tartozik-e vagy nem. Ezzel egy szegmentációs problémát oldunk így meg.

Ahhoz, hogy a későbbi feldolgozást egyszerűbb tegyük, az elszórt pontokból álló részeket megpróbálhatjuk téglalapokkal közelíteni. Ez a későbbiekre nézve azért előnyös, mert így ugyanazon algoritmus fogja tudni felismerni a szöveget, függetlenül attól, hogy az a bejegyzésnek melyik részéről származott. A különféle méretű és arányú téglalapok természetesen bonyodalmakat okozhatnak az implementációban. A problémát konkrét mérések segítségével lehet biztosan megoldani, de valószínűsíthető, hogy a nagyobb téglalapok megfelelő darabolása megoldást jelent rá.

4.5. Kontextus figyelembevétele

A rendszer létrehozásához szükséges kutatás egy igen érdekes része az, hogy a bejegyzések kontextusában lévő információkat hogyan lehet felhasználni a becslés pontosságának a javítása érdekében. A kontextushoz több összetevő is tartozhat, mint például az alábbiak.

- Az oldalon szereplő többi bejegyzés fontos információkat adhat az aktuálisan elemzett bejegyzésre vonatkozóan. Például, hogy ha a bejegyzés dátuma nehezen olvasható, akkor a szomszédos bejegyzések dátumai alapján már egy jó közelítő becslést adhatunk.
- A könyvekben lévő bejegyzéseket gyakran egy személy írta be huzamosabb ideig. A gépi tanulás alkalmassá tehető arra, hogy a személyhez kötődő jellegzetes vonásokat megtanulja, az előforduló betűméretek és betűformák alapján következtethessen olyan adatokra, amelyek az adott bejegyzésből egyébként nehezen, vagy egyáltalán nem lennének kinyerhetők.
- Az egyes települések anyakönyvezési szokásai mutathatnak kisebb eltéréseket. Amennyiben ismerjük ezeket, vagy a gépet minták alapján betanítjuk ezekre, úgy a sajátosságokban rejlő plusz információk is felhasználhatók.
- Az anyakönyvezési szokások időben is változtak. A szövegek felismerésénél a tollak típusa, a tinta színe is befolyásoló tényező. Például, hogy ha tudjuk, hogy időszakonként a bejegyzések barna tintával íródtak, akkor eleve az algoritmus preferálhatja az olyan vonalakat, amelyek ilyen színűek.
- A bejegyzéshez alakulásában a hatályos jogszabályok is közrejátszottak; például, a törvényi előírás bizonyos időszakban változott, akkor feltételekhez lehet kötni, egy-egy szöveges elemet hogyan próbáljon megtalálni a képelemző módszer.

4.6. Természetes nyelvi elemzés

A képek feldolgozása a klasszikus értelemben addig tart, amíg az adatok valamilyen szöveges formába nem kerülnek. Ez viszont a legritkább esetben jelenti azt, hogy a feldolgozásnak már közvetlenül használható formában megvan a végeredménye. Az adatok itt még csak törmelékesen, koordinátákhoz kötött karakterláncok alakjában adóttak. Ahhoz, hogy ebből olyan minőségű adatok legyenek, amelyek már átemelhetőek az elektronikus anyakönyvbe, egy további nyelvi elemzési lépésre is szükség van. Példaként az alábbi, szintén gépi tanulási módszerrel megoldható problémákkal kell még foglalkozni.

- A felismert szövegrészek egymáshoz közel, de nem egybefüggő szöveggé kerültek meghatározásra. Például a "január" hónap név az egymáshoz közeli "ja", "nu", "á" és "r" karaktersorozatokat tartalmazó szövegrészekben van. A természetes nyelvi elemző algoritmusnak meg kell próbálni ezekből egy értelmes szót összerakni.

- A fordított probléma is előfordulhat, vagyis amikor különálló szövegrészek kerültek összevonásra. Tipikusan az ilyen esetek szóközökkel elválasztott szavak esetében fordulnak elő. Például összevonásra kerülhet valakinek a két utóneve. Ekkor azt kell megvizsgálni, hogy névként előfordulhat-e az összevont eset, és hogy ha nem, akkor hol lehet azt elválasztani úgy, hogy értelmes legyen mindkét rész.
- A bejegyzésekben a sorok száma nem kötött. Az anyakönyvvezetőre volt bízva, hogy a bejegyzésnél hogyan tördelte a sort annak érdekében, hogy áttekinthető és olvasható legyen majd az eredmény. Ilyenkor egymás alatt is elhelyezkednek az összefüggő szövegrészek, amire szintén fel kell készíteni a feldolgozási módszert.
- Főleg a megjegyzés rovatba kerülhetnek be hosszabb, szabadszöveges részek. Ezeknél a természetes nyelvi elemzés különösen fontos és egyben komplikált is. Azt szeretnénk ugyanis elérni, hogy a szövegrész tördelésétől függetlenül kivehető legyen, hogy mi volt a szöveges tartalom, az hogyan tagolható mondatokra.
- Az elrendezés mellett a karakterfelismerési hibák is problémát jelentenek. A felismert szövegrésznél ellenőrizni kell, hogy abban előfordulhatott-e ilyen jellegű hiba. Például a nevek esetében kiindulhatunk abból, hogy a vezeték és keresztnév listája helyelközzel ismert. Amennyiben egy felismert szövegrésről tudjuk, hogy az egy név része, akkor célszerű azt megvizsgálni, hogy az említett listákban előfordul-e. Ez, ha nem vezet eredményre, akkor valamilyen szöveg metrika szerint meg lehet keresni a hozzá legközelebb eső neveket.

Az ilyen jellegű kutatások a természetes nyelvfeldolgozás (*Natural Language Processing*) tárgykörébe tartoznak (*Young et al. 2018*). A koordináták felhasználhatóak még ezen a szinten, mivel többletinformációkat tartalmaznak, viszont itt már nem kell, hogy a digitalizált szöveges részek kép formájában elérhetőek legyenek. Az alkalmazható módszertan megnevezése a *post OCR correction*, amely statisztikai elemzéseket és mesterséges neurális hálókat is használhat a minél pontosabb korrekciók érdekében (*Amrhein–Clematide 2018*).

4.7. A felismerési módszer automatikus hangolása

A felvázolt rendszer esetében a rendelkezésre álló tanítóminták száma a rendszer használata közben folyamatosan nő. Amikor egy anyakönyvvezető jóváhagy vagy javít egy gép által felismert szövegrészt, akkor azzal a gépi tanulási módszer pontossága növelhető. (Sajnos itt nem garantált közvetlenül az, hogy az új elemzendő bejegyzésekre biztosan jobban

fog működni a módszer. Annyi jelenthető csak ki, hogy a nagyobb mintaszám miatt nagyobb valószínűséggel fog tud pontosabb becslést adni az algoritmus.) Külön kutatás tárgyát képezheti az is, hogy milyen feltételek mellett, milyen módon, és milyen rendszerességgel érdemes a megnövekedett mintaszámra újra betanítani a módszert. Elképzelhető természetesen olyan változat is, amely a hangolást minden új minta után elvégzi, de az nagyon számításigényes lenne, és a nagy felhasználószám miatt sem tanácsos felügyelet nélkül hagyni egy statisztikai alapon működő rendszert.

A rendszer automatikus hangolása kapcsán érdemes arra felhívni a figyelmet, hogy mivel központosított rendszerről van szó, ezért az adott Anyakönyvi Hivatalban nyert tapasztalatokat az összes többi hivatalban is alkalmazni tudja a rendszer.

5. A kivitelezés feltételei és folyamata

A dolgozatban egy olyan rendszer felvázolására került sor az eddigiekben alapján, amely a kor modern gépi tanulási eszközkészletét feltételezi. A megvalósításra és a későbbi előnyökre vonatkozóan jó példát mutat az írek által készített adatkezelő rendszer (*Debruyne et al. 2016*). A kutatók a cikkükben részletesen kitérnek arra, hogy a szemantikus formában elérhetővé vált adatok milyen struktúrában állnak elő, azokon milyen módon és eszközökkel végezhetőek el további vizsgálatok.

A következő szakaszokban annak a részletezése következik, hogy a futurisztikusnak ható célok ellenére egy belátható időn belül megvalósítható és használatba vehető alkalmazási lehetőségről van szó.

5.1. Hardver eszközök

A megvalósítás kapcsán az alábbi hardverelemekre kell gondolnunk.

- A digitalizáláshoz szükség van egy megfelelő felbontással és képminőséggel rendelkező kamerára. Ezt egy olyan állványra kell majd tudni felszerelni, amely merőlegesen, minimális lencsetorzítás mellett képes felvételt készíteni az anyakönyvekről. A fényviszonyok biztosításához valamilyen fehér ledes fényforrás vagy derítő megfelelő lehet.
- Nyilvánvalóan nagyon nagy mennyiségű kép tárolására van szükség. Napjainkban már több terabyte méretű winchester-ek is

elérhető, így ez lokális vagy központosított tárolás esetén sem okozhat gondot.

- A hozzáférés biztosításához kell, hogy tartozzon a rendszerhez legalább egy központi szerver, amelyről az anyakönyvek digitalizált változatai elérhetőek. Ennél elsősorban a tárigény lehet a szempont, a számításigénye a koncepció szerint lényegesen nem különbözik bármely más tartalomkezelő rendszerétől.
- A gépi tanulási módszer finomhangolásához szükség van nagy teljesítményű számítások elvégzésére alkalmas számítógépekre. (Ezekről is kijelenthető, hogy egy részük bolti forgalomban kapható, de az egyetemek és kutatóintézetek többségében találhatóak ilyenek.)
- A rendszerhez hozzátartoznak a kliens számítógépek is. Azokat adottnak tekinthetjük, mivel már az Elektronikus Anyakönyvi rendszer kiépítésre került.

5.2. Szoftver eszközök

A rendszer működéséhez létre kell hozni egy olyan szoftver rendszert, amely a digitalizált képeket megoszthatóvá teszi, illetve a gépi tanulás elemzéseit el tudja végezni. A kivitelezési folyamat elsősorban ennek az elkészítését jelenti.

A gépi tanulásos, és általában a mesterséges intelligenciát alkalmazó szoftverekre is igaz, hogy az algoritmusok, szoftverkönyvtárak nyílt forráskódú alkalmazásként elérhetőek. A rendszer elkészítéséhez tehát ezek könnyen beszerezhetőek és felhasználhatók.

A gépi tanulás megvalósítása a Python programozási nyelven a *Scikit Learn* segítségével valósítható meg (*Hackeling 2017*). (A mesterséges intelligencia kutatások többségében ezen programozási nyelv és függvénykönyvtár használatával zajlanak.) Szintén használhatóak még a *Tensorflow* (*Abadi et al. 2016*), illetve a *PyTorch* (*Paszke et al. 2017*) nevű könyvtárak is.

5.3. Kutatás és fejlesztés

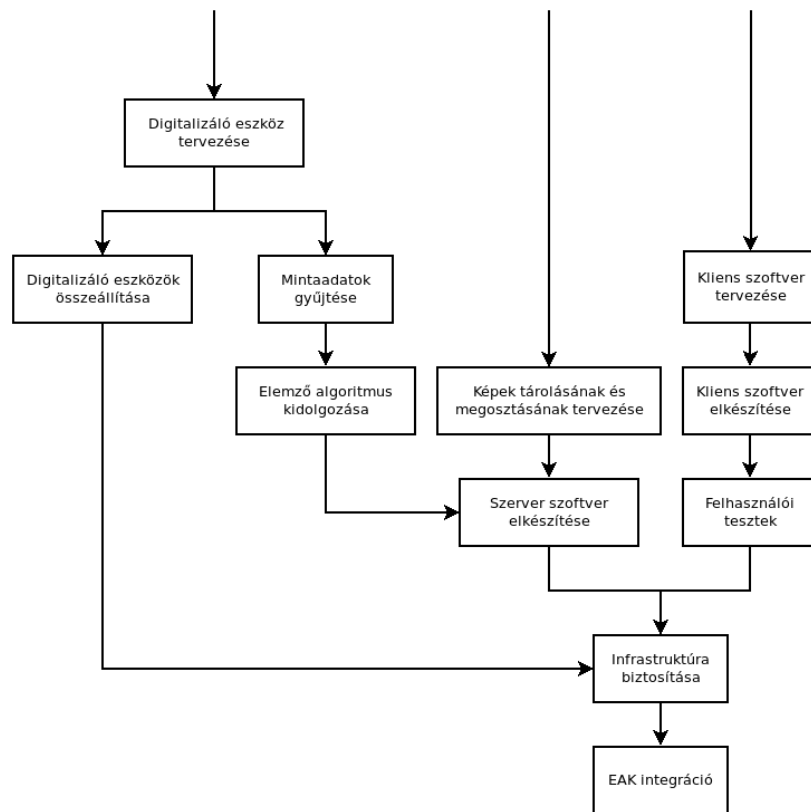
A rendszer részleteinek megtervezéséhez, a konkrét algoritmusok kidolgozásához, a működés validálásához és a rendszer további elemeinek a megvalósításához szükség van kutatókra és fejlesztőkre. Mivel a problémakör a jel-, azon belül is a képfeldolgozáshoz tartozik, ezért napjainkban népszerű, sok helyen oktatják, és ipari szinten használják a módszereit. A kutatási és fejlesztési fázis eredményeképpen kell elkészülnie egy olyan prototípusnak, amellyel már ki lehet próbálni, hogy a tervek mennyire működőképesek az Anyakönyvi Hivatalokban.

5.4. Humán erőforrások

Az anyakönyvvezetők, kutatók, fejlesztők, üzemeltetéssel foglalkozó szakemberek mellett szükség van még olyan emberekre, akik az anyakönyvek digitalizálását végzik. A digitalizáló eszköz használata egyszerű, így könnyű fizikai munkaként szinte bárki el tudja végezni.

5.5. A tervezett kivitelezési folyamat

A szükséges háttér áttekintését követően most azt vizsgáljuk meg, hogy milyen lépésekben végezhető el a rendszer kivitelezése. A kivitelezési folyamat áttekintését láthatjuk a 2. ábrán.



2. ábra: A megvalósítási folyamat áttekintése

Forrás: saját szerkesztés

A kivitelezés egyaránt kezdődhet a digitalizáló eszköz tervezésével, a képek tárolásáért és megosztásáért felelős alrendszer tervezésével vagy pedig a kliensszoftver kialakításával.

Amennyiben a digitalizáló eszköz már elérhető, akkor lehetőség van mintaadatok gyűjtésére, amelyek már a gépi tanulós módszer kutatásához, hangolásához megfelelőek lesznek. A kivitelezés vélhetően legbonyolultabb és így legtöbb ideig tartó része az elemző algoritmus kidolgozása. Ez fogja majd tudni megmutatni, hogy egyáltalán mennyire jól lesz használható a rendszer a tervezett feladatok ellátására.

A kliens szoftver elkészítése külön ágon folyhat, mivel az nem feltételezi sem az eszköz rendelkezésre állását, sem pedig a képek megosztását. (Ilyen esetben a kapcsolódó felületekhez ideiglenes, úgynevezett *mock* szerverek hozhatók létre, hogy a funkcionalitást tesztelni lehessen.)

A szerver szoftverének elkészítése azt jelenti, hogy a rendszer szoftver szempontjából már telepítésre és tesztelésre kész állapotban van. A kliensre vonatkozó felhasználói tesztek azért szükségesek, hogy még az éles üzembehelyezés előtt minél több használhatóságra vonatkozó problémára fény derülhessen.

Országos szinten nagyon sok anyakönyv digitalizálására van szükség. A digitalizáló eszközök összeállításához nyilvánvalóan be kell szerezni az eszközöket, esetlegesen fel kell keresni szakembereket is, hogy az eszközök nagy mennyiségben elkészíthetők legyenek.

Az infrastruktúra biztosítása azt jelenti, hogy hardver, szoftver és üzemeltetés tekintetében is működésre készen áll a rendszer.

Az utolsó lépésként szereplő EAK integráció tulajdonképpen az éles üzembehelyezést jelenti, vagyis amikor már az anyakönyvvezetők használatba veszik a rendszert.

6. Összegzés

A dolgozat egy koncepciót vázolt, amellyel kiválthatók a jelenleg eredeti papír formájukban használt anyakönyvek. A bennük szereplő adatok digitalizálási módjára tesz javaslatot, és arra, hogy a gépi tanulást alkalmazó karakter- és szövegfelismerő alkalmazások segítségével hogyan emelhetők át a bejegyzések az aktuálisan már használt elektronikus rendszerekbe.

A mesterséges intelligencia eszközei elérhetőek, a probléma jól definiált, viszont a módszer megvalósítása további kutatásokat igényel. Ez minden hasonló terv esetében csak így történhet, mivel a gépi tanulási módszerek mintákból tanulnak, így a részproblémák, esetleges nehézségek a kutatás fázisában tudnak kiderülni.

A dolgozat igyekezett a lehető legnagyobb részletességgel leírni a kivitelezés módját. Ez azért tűnt fontosnak, mert a mesterséges intelligencia megnevezés alatt gyakran a földtől kissé elrugaszkodott, nehezen realizálható alkalmazási módokat lehet érteni. Jelen esetben viszont egy olyan alkalmazási lehetőséggel állunk szemben, amelyik a szükséges humán- és tárgyi feltételeiben könnyen elérhető, néhány éves időtávban kivitelezhető. Remélhetőleg a dolgozat eléri célját, és a felvetett ötletek megvalósításra kerülnek majd, segítve ezzel az anyakönyvvezetők munkáját.

Irodalomjegyzék

ABADI, M., B. ARHAM, P., CHEN, J., CHEN, Z., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., IRVING, G., ISARD, M., ET AL. (2016) Tensorflow: A system for large-scale machine learning. In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pp. 265–283.

AL-KARMI, A. N., SINGH, S. S.–SOOR, B. S. (1999) *Optical character recognition of handwritten or cursive text*. Jan. 19. 1999. US Patent 5,862,251.

AMRHEIN, C., CLEMATIDE, S. (2018) Supervised OCR error detection and correction using statistical and neural machine translation methods. *Journal for Language Technology and Computational Linguistics (JLCL)*, Vol. 33. No. 1. pp. 49–76.

BAL, A., SAHA, R. (2016) An improved method for handwritten document analysis using segmentation, baseline recognition and writing pressure detection. *Procedia Computer Science*, Vol. 93. pp. 403–415.

BRADSKI, G., KAEHLER, A. (2008) *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc.

DEBRUYNE, C., BEYAN, O. D., GRANT, R., COLLINS, S., DECKER, S.–HARROWER, N. (2016) A semantic architecture for preserving and interpreting the information contained in irish historical vital records. *International Journal on Digital Libraries*, Vol. 17. No. 3. pp. 159–174.

HACKELING, G. (2017) *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd.

PASZKE, A., GROSS, S., CHINTALA, S., CHANAN, G., YANG, E., DEVITO, Z., LIN, Z., DESMAISON, A., ANTIGA, L.–LERER, A. (2017) Automatic differentiation in pytorch. *31st Conference on Neural Information Processing Systems*.

PERRY, S. R. (2014) Digitization and digital preservation: A review of the literature. *School of Information Student Research Journal*, Vol. 4. No. 1. pp. 4.

RODRIGUES, R. J., THOMÉ, A. C. G. (2000) Cursive character recognition-a character segmentation method using projection profile-based technique. In: *The 4th World Multiconference on Systemics, Cybernetics and Informatics SCI 2000 and The 6th International Conference on Information Systems, Analysis and Synthesis (ISAS)*.

SCHOMAKER, L. (2016) Design considerations for a large-scale image-based text search engine in historical manuscript collections. *it-Information Technology*, Vol. 58. No. 2. pp. 80–88.

SCHREIBER, S., AGNE, S., WOLF, I., DENGEL, A.–AHMED, S. (2017) Deepdesrt: Deep learning for detection and structure recognition of tables in document images. In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 1. IEEE. pp. 1162–1167.

SILVA, J. C., HIGGINS, C. T. (2019) *License plate distributed review systems and methods*, Mar. 19. 2019. US Patent App. 10/235,332.

STARR, P., STARR, S. (1995) Reinventing vital statistics. the impact of changes in information technology, welfare policy, and health care. *Public health reports*, Vol. 110. No. 5. pp. 534.

YILDIZ, M. (2007) E-government research: Reviewing the literature, limitations, and ways forward. *Government information quarterly*, Vol. 24. No. 3. pp. 646–665.

YOUNG, T., HAZARIKA, D., PORIA, S.–CAMBRIA, E. (2018) Recent trends in deep learning based natural language processing. *IEEE Computational intelligence magazine*, Vol. 13. No. 3. pp. 55–75.