

Horváth Iván

Kulturális
Örökség
Gép

Cultural
Heritage
Machine

Q.E.D.
Budapest, 2023

Egy 2004-ben született javaslat

Régen a könyvolvasáshoz az embernek könyvet kellett a kezében tartania. Nagyon sok, lényegében egyidejű olvasási esemény előidézéséhez nagyon sok könyvpéldányra volt szükség. Ma ugyanannak a kiszolgáló számítógépnek ugyanazt a tárolóeszközét számtalan távoli felhasználó lényegében egyidejűleg igénybe veheti.

A szerző már 20 évvel ezelőtt alapvető problémára hívta fel a figyelmünket ezzel kapcsolatban: az internet megjelenésével a szövegszorosítás energiaigénye tovább csökken, de ez nem jár együtt a szöveghordozó példányok megsokszorozásával. A hozzáférés számossága többé már nem függvénye a példányszámnak, s ez eltakarja előlünk azt a problémát, hogy jelenlegi formájában az emberiség történetének legsérülékenyebb kulturális örökségét állítjuk elő. Az internetet könyvtárként kezeljük, pedig nem feladata a megőrzés. Nem könyvtár, hanem pályaudvar.

Ez az oka annak, hogy – szemben a történelem általános tapasztalatával, amikor a sokszorosított hordozópéldányok megsemmisülése általában nem járt együtt a szövegek megsemmisülésével – az internet korában a példányokkal együtt gyakran maguk a szövegek is elpusztulnak.

Pedig az emberiség teljes digitális öröksége – kb. 70 kultúrával számolva – tulajdonképpen csekély méretű, így sokmilliószorosítás révén megőrizhető.

A kiadó

A Proposal from 2004

In the past, for one to read a book, one had to hold a physical copy of that book in their hand. For a large number of readers to have an essentially simultaneous reading experience of the same book, an equally large number of physical copies of that book had to be available. Today, a single storage device on a single server computer can be accessed by a large number of remote users at essentially the same time.

Already two decades ago, the author highlighted a pivotal issue in this context: although the advent of the Internet has further decreased the energy required for text reproduction, it has not contributed to the multiplication of the number of copies of the actual text media. The number of views is no longer a function of the number of copies available. This obscures the problem that we are producing what—in its current form—is the most fragile body of cultural heritage in human history. We treat the Internet as if it were a library, even though it has never been meant to preserve our cultural heritage. It is not a library; it is a railway station.

Thus, while historically the destruction of physical reproductions of a text has not typically resulted in the loss of the text itself, in the age of the Internet, the destruction of the medium often leads to the loss of the text as well.

A lamentable circumstance, really, considering that the entire digital heritage of humankind, encompassing approximately 70 cultures, is a relatively small body of data that could be preserved by reproducing it in a few millions of copies.

Ez a mű nem köztulajdon, hanem a GNU General Public Licence (GPL © 1991) hatálya alá tartozó szellemi termék. Ezért a következő megjegyzés szövegétől elválasztani nem szabad.

Kulturális Örökség Gép. Copyright © 2004–2023 Horváth Iván. Mindenkinek joga van arra, hogy ezt a művet egészében vagy részben, bármely tetszőleges eszköz igénybevételével, akár papírra, akár más adathordozóra lemásolja, ott korlátozás-mentesen tárolja, ingyenesen terjessze, vagy pénzért árusítsa, feltéve, hogy ezt a címléírást, ezt a copyright-ot és ezt az engedélyező megjegyzést minden részleges vagy teljes másolati példányon jól olvasható módon feltünteti. Engedélyezetlen módosított változatokat nem szabad készíteni. Fordítások készítése nem számít módosításnak.

This work is not in the public domain but is an intellectual product released under the terms of the GNU General Public Licence (GPL © 1991). It should therefore not be separated from the text of the following note.

Cultural Heritage Machine. Copyright © 2004–2023 by Iván Horváth. Everyone has the right to copy, either in whole or in part, using any means available, whether on paper or any other storage medium, to store on such medium without any restriction, distribute freely, or sell for profit this work, provided that this citation, copyright notice, and licensing statement are prominently displayed in a legible manner on all partial or complete copies. No unauthorized modifications or altered versions are permitted. Translating is not considered an alteration.

Kulcsszavak: kulturális örökség, megőrzés, sokszorosítás

Keywords: cultural heritage, preservation, reproduction

Tartalom

Kulturális Örökség Gép 7

Cultural Heritage Machine 23

Kulturális
Örökség
Gép



Az eljárás az UNESCO 2003. október 15-én kibocsátott *Charter on the Preservation of Digital Heritage* című felhívásának 5. cikkelyében felvetett problémára ajánl megoldást. A digitális dokumentumok megőrzésének problémáját a legtöbb kutató adatóvóhelyek építésével és az ezekben elhelyezett tárolóeszközök műszaki megbízhatósági fokának emelésével oldaná meg. A találmány ezzel szemben a tárolóeszközök mennyiségét javasolja radikálisan megnövelni. A tárolóeszközök számát valószínűleg csak akkor lehet gazdaságosan felemelni a szükséges mértékre, ha pedagógiai segédeszközökként, tanszerekként is használni lehet őket.

1. A találmány tárgya

1.1 Eljárás a kulturális örökség digitalizált vagy eleve digitális dokumentumainak hosszú távú megőrzésére.

1.2 Eljárás a kulturális örökség digitalizált vagy eleve digitális dokumentumai pedagógiai közvetítésének elősegítésére.

1.3 Az 1.1 és 1.2 pontban megnevezett eljárás összekapcsolása.

2. A technika állása

2.1 Általános vélemény – például az *International Technology Roadmap for Semiconductors*¹ áttanulmányozásából erre következtethetünk –, hogy a digitális tárolók kapacitásbővülésének, továbbá méretük, energiaigényük és áruk viszonylagos csökkenésének folyamata a következő évtizedekben aligha fog lelassulni. A találmány műszaki feltétele az, hogy a belátható jövőben tömegesen, viszonylag olcsón és viszonylag kis méretben rendelkezésre álljanak 1 petabájtnál nagyobb, később pedig esetleg lényegesen nagyobb tárolóegységek.² Ha ez a feltétel nem teljesül, a találmány nem valósítható meg.

2.2 A digitalizált dokumentumok fennmaradását nehezítő tényező a számítógépek és a programok gyors elavulása. Az utóbbi évtizedekben megjelentek és az interneten gyorsan terjednek a gépfüggetlen programozási³ és dokumentumleíró⁴ nyelvek. Ha ez a folyamat visszajára fordul, a találmány nem valósítható meg.

1 <http://public.itrs.net/> – *A szöveg megfogalmazása (2004) óta elhalt hivatkozás (2023. 11. 10.)*

2 Az alábbiakban ismertetendő eljárásnak része egy olyan be rendezés is (CHM), amely ma még nem létezik. Az ennek bemutatásában olvasható számok csupán tájékoztató jellegű becslések.

3 Például a JAVA: <http://java.sun.com/>

4 Például az SGML és leszármazott nyelvei (HTML, XML stb.). Vö.: <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16387>

3. A megoldandó feladat

Digitalizálás révén a kulturális dokumentumok rendkívül könnyen hozzáférhetővé tehetők a kutatás, oktatás és művelődés számára. A digitalizált kulturális dokumentumok hátránya, hogy roppant sérülékenyek, olyannyira, hogy a *European Commission on Preservation and Access* főtitkára⁵ szerint az archiválás egyoldalú digitális technikái mellett tulajdonképpen továbbra is fenn kellene tartani az analóg másolatok készítésének eljárásait is,⁶ jóllehet így le kell mondanunk a dokumentumok széleskörű hozzáférhetővé tételéről.

Ugyanennek az általános problémának orvoslását sürgeti a *Digitális Örökség Megőrzésének Chartája*,⁷ amelyet az UNESCO 2003. október 15-én fogadott el. A *Charta* 1. cikkelye a digitális örökséget állandóan növekvő kincsként írja le, amely részint a korábban nem digitális formában létező kulturális értékek digitalizálása nyomán képződik, részint

5 <http://www.knaw.nl/ecpa/members.html#staff> – A szöveg megfogalmazása (2004) óta elhalt hivatkozás (2023. 11. 10.)

6 Yola de Lusenet, *Keeping Things That Work. Preservation Aspects of Digitization*, in Andrea Bozzi, Laura Cignoni, Jean-Louis Lebrave (eds.), *Digital Technology and Philological Disciplines*, Pisa–Roma, Istituti Editoriali e Poligrafici Internazionali, 2004, 145–161.

7 http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC &URL_SECTION=201.html – A szöveg megfogalmazása (2004) óta elhalt hivatkozás (2023. 11. 10.)

eleve digitális formában jön létre.⁸ A 3. és 4. cikkely megállapítja, hogy a gépek és programok gyors elavulása és egyéb tényezők folytán a digitális örökség rohamosan pusztul. A pusztulás „gyors és elkerülhetetlen”, ha a tagállamok nem találnak sürgősen ellenszert. Az 5. cikkely fogalmazza meg azt a feladatot, amelynek megoldásához való hozzájárulás a találmány célja: „Alapvető fontosságú a digitális örökség folytonosságának fenntartása. A digitális örökség megőrzéséhez a digitális információ teljes életciklusát átfogó intézkedéssorozatra van szükség, az információ keletkezésétől a hozzáférésig. A digitális örökség hosszú távú megőrzése azzal kezdődik, hogy olyan megbízható rendszereket és eljárásokat

8 “The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are “born digital”, there is no other format but the digital object.

Digital materials include texts, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained. Many of these resources have lasting value and significance, and therefore constitute a heritage that should be protected and preserved for current and future generations. This ever-growing heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression.”

kell kidolgozni, amelyek hiteles és tartós digitális tárgyakat állítanak elő.”⁹

A találmány a digitális örökség dokumentumainak tartós fennmaradását célozza.

4. A feladat általános megoldása

4.1 Fogalmak

4.1.1 *Kulturális örökség* az emberiség teljes felhalmozott – anyagi és szellemi – művelődési kincse.¹⁰

4.1.2 A *kánon* a kulturális örökség kicsinyített képe: a tankönyvek műveltséganyaga.

4.1.3 *Dokumentum* bármely szöveg. Szöveg az is, ami eredetileg nem szöveg volt, hanem képből, mozgóképből, hangból vagy egyéb tárgyból való átalakítás (például digitalizálás) révén jött létre.

4.1.4 A *szöveg* jelek rendezett halmaza.

4.1.5 *Jel* az, amit egy jelkészlet-halmaz egy elemébe leképezünk.

9 “Continuity of the digital heritage is fundamental. To preserve digital heritage, measures will need to be taken throughout the digital information life cycle, from creation to access. Long-term preservation of digital heritage begins with the design of reliable systems and procedures which will produce authentic and stable digital objects.”

10 http://portal.unesco.org/culture/en/ev.php-URL_ID=2185&URL_DO=DO-TOPIC&URL-SECTION=201.html
– A szöveg megfogalmazása (2004) óta elhalt hivatkozás (2023. 11. 10.)

4.1.6 *Pontos másolás* történik akkor, amikor a forrásul szolgáló szöveg elemeit és az eredményül kapott szöveg elemeit a jelkészlet-halmaznak rendre ugyanazokba az elemeibe képezzük le.

4.2 Kifejtés

Az utolsó meghatározás értelmében csak szöveget lehet pontosan lemásolni. Amit nem lehet lemásolni belőle, az nem is része a szövegnek. A szövegdigitalizálás csupán a pontos másolás egy fajtája. Képek és hangok digitalizálása pedig nem egyéb, mint szöveggé alakításuk. Csak szöveges adatbázisok vannak: nemcsak a metaadatok, hanem a kép- és hangadatok is csak már szövegekké alakulva kerülhetnek be az adatbázisba.

A számítógépek korát ebben az értelemben nem a szövegek visszaszorulása, az úgynevezett „Gutenberg-galaxis” vége¹¹ jellemzi, hanem sokkal inkább a szövegeken alapuló kommunikáció uralkodóvá válása.

A látható és hallható felszín gyakran kép, mozgókép, hang – a mélyben azonban szövegek áramlanak.

A történelem során a szövegekhez való hozzáférés növekszik, miközben a szöveghordozó tartóssága csökken.¹²

11 *Kép, beszéd, írás*, szerk. Neumer Katalin, 2003.

12 A hordozó tartóssága általában gyengül: sziklafelirat, kiégett agyagtábla, papirusz, pergamen, az ősnymtatványok papírja, rotációs papír, optikai és mágneses jelhordozók.

Ennek az általános (kivételeket is megengedő) történeti szabálynak az a magyarázata, hogy a szövegekhez való hozzáférést egyre inkább a szövegek sokszorosítása (több példányban való pontos lemásolása) révén érik el.

Már az ókorban felismerték, hogy a kevésbé ellenálló anyagból készült szöveghordozó példányok megsokszorozása a szövegeknek nemcsak térbeli elterjedését, hanem időbeli fennmaradását is jobban elősegíti, mint a szöveg ércbe öntése: Horatius szerint a papirusztekercsre írt remekmű „aere perennius”, ércnél maradóbb.¹³ Igaza volt: Cicero *De amicitia* és *De officiis* című műveiből ma például darabonként több mint 600 kéziratos forrást tartunk nyilván.¹⁴ A szöveg fennmaradásához a szöveghordozó példánynak csupán addig kell kitartania, amíg a szöveget le nem másolják. Mennél kevesebb energia kell a lemásoláshoz, annál nagyobb a bekövetkezés valószínűsége. Az iparszerű gyakorlatban is ugyanígy: mennél kevesebb energiát kell közölni a szöveghordozó egy-egy példányával, annál könnyebb megoldani a sokszorosítás feladatát. Ezért terjedt el a könyvnyomtatás.

A sokszorosítás révén való szövegmegőrzés módszerének azonban hátránya is van, nemcsak előnye. Az a szöveghordozó, amellyel a sokszorosítás során kevesebb energiát közöltek, azért kevésbé tartós,

13 HORATIUS Flaccus, Quintus, III, XXX, i.

14 Paolo Fedeli, *Le trappole della contaminazione*, in *Digital Technology and Philological Disciplines*, Pisa–Roma, Istituti Editoriali e Poligrafici Internazionali, 2004, 213.

mert elveszíteni is kevesebb energia hatására fogja szöveghordozó képességét. Az alacsony energiaigényű sokszorosítás csak akkor tartja fenn a szövegeket, ha valóban rendszeresen be is következik. A sokszorosítás által való szövegmegőrzés a társadalom részéről állandó tevékenységet feltételez.

Ez a magyarázata annak a történeti szabálynak, hogy a szövegekhez való hozzáférés és a szövegek fennmaradási esélye annak ellenére javul, hogy közben a szöveghordozó tartóssága romlik. A szövegek világát állandó másolási tevékenységgel fenntartó társadalomnak köszönhető, hogy a szövegek túléltek szerzőjüket, sőt a nyelvet magát is, amelyen íródtak.¹⁵

Az internet nem teljesen illik bele a történeti szabályba. A szövegsokszorosítás energiaigénye ugyan tovább csökken, de nem jár együtt a szöveghordozó példányok megsokszorozásával. Régen a könyvolvasáshoz az embernek könyvet kellett a kezében tartania. Nagyon sok, lényegében egyidejű olvasási esemény előidézéséhez nagyon sok könyvpéldányra volt szükség. Ma ugyanannak a kiszolgáló számítógépnek ugyanazt a tárolóeszközét számtalan távoli felhasználó lényegében egyidejűleg igénybe veheti. Anélkül, hogy a részletekbe belemennénk, megállapíthatjuk, hogy az internetes szövegáramlás során

15 Iván Horváth, *On-line Critical Editions in the 90s: Practice and Theory*, in Andrea Bozzi, Laura Cignoni, Jean-Louis Lebrave (eds.), *Digital Technology and Philological Disciplines*, Pisa–Roma, Istituti Editoriali e Poligrafici Internazionali, 2004, 245–258.

a vevőoldalon olykor nem jelenik meg teljes értékű példány. A vevőoldali példány szelektív (épp csak a megkívánt információt tartalmazza), töredékes (nem tartalmaz információt sem a szöveggörnyezetéről, sem a szöveggörnyezet tágabb szerkezetéről), és olykor nem is kerül megőrzésre. Teljes értékű példányoknak csak azok tekinthetők, amelyek az adóoldalon helyezkednek el. Ilyen a kiszolgáló számítógép és a tükörszerver tárolóeszköze, valamint a mentések – vagyis roppant kevés és sérülékeny példány. Ez az oka annak, hogy – szemben a történelem általános tapasztalatával, amikor a sokszorosított hordozópéldányok megsemmisülése általában nem járt együtt a szövegek megsemmisülésével – az internet korában a példányokkal együtt gyakran maguk a szövegek is elpusztulnak.¹⁶

A történeti szabály szerint a szövegekhez való hozzáférés növekszik. Ez azt jelenti, hogy a kulturális örökség dokumentumai közül azoknak a jelentősége óhatatlanul lecsökken, amelyek nem lesznek elérhetők az internet hálózatán keresztül. Azok viszont, amelyek elérhetők lesznek az internet hálózatán keresztül, elpusztulnak, mert a történelemből ismerünk olyan régi szövegeket, amelyek kevés példányban léteztek, de tartós hordozón, ismerünk olyan régi szövegeket is, amelyeket gyenge hordozókra írtak, de sok példányban, nem ismerünk vi-

16 Iván Horváth, *Magyarok Babelben*, Szeged, JATEPress, 2000, 131–135. On the exhibition at the Frankfurt Book Fare of texts that have disappeared from the Internet during the nineties, see <http://magyar-irodalom.elte.hu/contentware/magyar/>

szont olyan régi szövegeket, amelyeket gyenge hordozókra, kevés példányban rögzítettek volna. Az internet a kulturális örökség digitalizált vagy eleve digitális dokumentumainak megőrzését nem teszi lehetővé.

A megoldásban két módszert is alkalmazhatunk: vagy a hordozó tartósságát fokozzuk, vagy a példányszámot növeljük.

Ha inkább az első módszert választjuk, akkor csökkenthetjük a digitális dokumentumok sérülékenységét, fokozhatjuk a hordozók működésének a megbízhatóságát, végezhetjük a kódolást nagyobb energiaráfordítással. Ez az, amire a nagy közgyűjtemények törekednek, és amire a mai kutatás elsősorban irányul. A módszer hátránya, hogy ellentmond a történeti szabálynak. Ez az elavult, óegyiptomi és óperzsa út: a piramis legmélyén elhelyezkedő sírkamra falára festett szövegnek, a megközelíthetetlen hegyoromra felvett sziklafeliratnak módszere.

A második módszer a példányok megsokszorozásáé. Előnye, hogy megvalósítja a történeti szabályt. Az interneten egymással kapcsolatban álló berendezéseket csupán kiegészíti egy újfajta részegységgel: a kulturális örökség nagy és növekvő részét hordozó, központilag írható és frissíthető, a felhasználók által pedig csak olvasható tároló néhányszor 10^6 – 10^9 példányával.

4.3 A találmány

Nagyon nagy (lehetőleg milliós-milliárdos nagyságrendű) sorozatban gyártott, központilag írható és frissíthető, a felhasználók által csak olvasható tároló, amely a kulturális örökség digitalizált vagy eleve digitális dokumentumait tartalmazza. Ideiglenes neve: Cultural Heritage Machine (CHM).

A CHM adott esetben a felhasználók szűkebb csoportja által is használható, de lehetőség szerint személyi használatra szánt eszköz. A következőkben a leírás az egyszemélyi felhasználást tételezi fel.

A CHM szabványos csatolóval ellátott, nagyon nagy kapacitású (legalább 1 petabájtos), mágneses, optikai vagy egyéb elven működő, írható és olvasható tárolóegység, amely nem igényel energiát ahhoz, hogy a benne rögzített adatokat néhány évig megőrizze. Működési energiaigénye, helyigénye, zaja, tömeggyártás esetén ára viszonylag alacsony. Ellátható ugyan saját műveleti egységgel és intelligens felhasználói felülettel,¹⁷ de alapvető felhasználási módja az, amikor szabványos csatolóval a felhasználó mindenkori saját személyes hálózati kommunikációs készülékéhez/készülékeihez kapcsolják hozzá. Forgalomba hozható más készülékekkel egybeépítve vagy önálló modulként. Különös megbízhatósági és tartóssági követelményeknek nem kell eleget tennie.

A CHM-ben tárolt adatok viszonylag rendezettek. Az adatok rendezéséhez, a közöttük való kereséshez ugyanolyan felhasználói felületet kell használni,

¹⁷ Erre különösen a régi adatbázis-kezelő programok emulált gépi környezetben való futtatásakor lehet szükség.

mint akkor kellene, ha a CHM tartalma a hálózaton keresztül lenne hozzáférhető. A CHM tartalma korlátozás nélkül használhat hálózati hiperhivatkozásokat.

A CHM tárolónak törzs- és kiegészítő része van. Törzsrészét illetően a felhasználó nem rendelkezik írási joggal. Írási joggal a társadalom által ezzel felruházott intézmény rendelkezik. Az intézmény vagy megbízottja végzi el az adatok frissítését is. Kívánatos, hogy a frissítések ne a korábbi anyag egyszerű felülírásai legyenek, hanem őrizzük meg a frissítések történeti egymásutánját; a frissítési rétegek maradjanak helyreállíthatók és vizsgálhatók.

A CHM tárolónak kiegészítő (nem törzs-) részét illetően csak a felhasználónak van írási joga. A kiegészítő rész fizikailag elhelyezhető a CHM-en kívül is. Felhasználói felülete ugyanolyan, mint a hálózaté és mint a CHM törzsrészéé.

A CHM törzsrészében elhelyezett dokumentumok három csoportra oszthatók. Ezek: a nemzeti, a nemzetközi, továbbá az általános kulturális örökség anyaga. Az első csoportban enyhe szűrést, a másodikban szigorú válogatást hajt végre az írásjoggal felruházott intézmény. Az, ami az egyik nemzet számára az első csoportba tartozik, az egy másik számára a másodikba – így a különböző nemzetekhez tartozó személyi CHM-ek anyaga jól kiegészíti egymást. Az általános (nemzethez nem sorolható) kulturális örökséget enyhe szűréssel kell válogatni.

A kulturális örökség dokumentumait a CHM számára érzékszervileg elfogadható, érzékszervileg hiteles, illetve az érzékszervi hitelességet messze

meghaladó részletfelbontással lehet digitalizálni. Kívánatos, hogy a történetileg javuló felbontóképességű digitalizálások a CHM tárolókapacitásának történeti növekedésével tartsanak lépést.

A CHM adattárolási és címzési módja nemzetközi szinten szabványos. Így megoldjuk, hogy a CHM-ek felhasználói a kulturális örökségre vonatkozó megfigyeléseiket jól érthető módon kicserélhessék egymással.

4.4 Pedagógiai segédeszköz

A találmány a 4.3 pontban leírt CHM mint pedagógiai segédeszköz, tanszer.

A kulturális örökség kicsiny része a tankönyvek műveltséganyaga, a kánon. A kánon terjedelme csökken, a kulturális örökség terjedelme nő.

A mai oktatáskutatók egy csoportja úgy gondolja, hogy a műveltségi kánont gyökeresen tovább kell csökkenteni, és a tanulóknak inkább a gyakorlati életben használható problémamegoldó képességeit kell fejleszteni. Egy másik csoport az elitképzésben helyesnek tartja a hagyományos műveltséganyag pozícióinak megőrzését.¹⁸ A CHM mindkét pedagógiai felfogásba beilleszthető mint tanszer.

Az elitképzésben részt vevő tanuló – majd pedig a felnőtt – természetesen tud mit kezdeni a rendelkezésre bocsátott kincscsel, amely gyakorlatilag

18 *Tartalmak és módszerek az ezredforduló iskolájában. Tanulmányok a tantárgyi helyzetfelmérésről*, szerk. KERBER Zoltán, 2004, különösen 45–67.

végtelen műveltségforrásnak tekinthető. Az elitképzésből kimaradó tanuló – majd pedig a felnőtt –, aki nem tudja vagy nem akarja személyisége részeként birtokolni a műveltséget, a CHM révén birtokolhatja azt olyan tárgyi tulajdonként, amellyel bánni tud.

Az iskola fő feladata mindkét esetben a CHM kezelési eljárásainak megtanítása, a kulturális örökséggel való bánás begyakorlása. A CHM a tanulótól minden esetben aktív közreműködést, a tanártól pedig merőben új pedagógiai módszertant igényel. Az a cél, hogy a CHM-et az ember gyermekkorban, esetleg ritualizált formában vegye át, és őrizzen meg belőle egy példányt egész életében; váljék a személyiség állandó kísérőjévé, mint a szemüveg vagy a mobiltelefon.

Megjegyzés. A CHM elterjesztésére a jogalkotásnak és az üzleti életnek egyaránt lehetnek eszközei. A CHM-ben tárolt dokumentumok jogi helyzetének tisztázására több javaslat is felhasználható.¹⁹

A 4.1.5 és 4.1.6 pont értelmében a dokumentumok értelmezéséhez szükséges az odavágó jelkészlet-halmaz birtoklása. A kulturális örökség megőrzésének csak akkor van értelme, ha az örökséget igénylő, annak befogadására kész ember is fennmarad. Ezért a 4.3 és 4.4 pontban leírt két feladat (a kulturális örökség megőrzése, illetve tanítása a CHM segítségével) feltételezi egymást.

19 Iván Horváth, *Művelődés közpénzen*, in HVG, 9 June 2001. Bálint Magyar, *Network of Clearinghouses for Educational Software*, Rome, January 2004 (presentation): <http://magyar-irodalom.elte.hu/biop/barbar/NCES.PDF>

Az iskolai oktatás mint tömeges felvevőpiac, lehetővé tenné a CHM gazdaságosan nagy sorozatban való gyártását. És megfordítva: a CHM nagy sorozatban való gyártása tenné lehetővé a kulturális örökség digitalizált vagy eleve digitális dokumentumainak fennmaradását.

5. Szabadalmi igénypontok

5.1 Eljárás a kulturális örökség digitalizált vagy eleve digitális dokumentumainak hosszú távú megőrzésére, azzal jellemezve, hogy nem a tárolóeszközök megbízhatósági fokát, hanem a mennyiségét növeljük meg.

5.2 Eljárás az 5.1 igénypontban meghatározott módszer alkalmazására a kulturális örökség digitalizált vagy eleve digitális dokumentumainak pedagógiai közvetítésében, azzal jellemezve, hogy az ott leírt tárolóeszközöket pedagógiai segédeszközként használjuk fel.

Cultural
Heritage
Machine

This process offers a solution to the problem described by Article 5 of the *Charter on the Preservation of Digital Heritage* adopted by UNESCO on 15 October 2003. Most researchers suggest solving the problem of the preservation of digital documents by constructing data shelters and by increasing the dependability of the storage devices placed in these data shelters. This invention would, instead, radically increase the quantity of the storage devices relied upon. The quantity of the storage devices can only be increased to the desired level economically if the devices can also be used as teaching resources.

1. The subject of the patent

1.1 A process for the long-term preservation of digital documents–information created digitally or converted into digital form from existing analogue resources–representing cultural heritage.

1.2 A process for the promotion of the pedagogical transmission of digital documents–information created digitally or converted into digital form from existing analogue resources–representing cultural heritage.

1.3 Connecting the processes named under 1.1 and 1.2 above.

2. The current state of technology

2.1 The general view—as reflected in, among others, the *International Technology Roadmap for Semiconductors*¹—is that the trend seen in the expansion of the capacity of digital storage devices, as well as in the reduction of their relative size, energy consumption, and price, is unlikely to slow down in the coming decades. This invention relies on the assumption that in the near future, mass-produced, relatively cheap, and relatively small storage devices with a storage capacity larger than—and eventually, significantly larger than—1 petabyte (PB) will become available.² If this condition is not fulfilled, the invention cannot be realized.

2.2 The preservation of digitised documents is hindered by the rapid obsolescence of computer hardware and software. Over the past few decades, there has been a rapid proliferation of device-independent programming³ languages and document mark-up languages.⁴ If this process is reversed, the invention cannot be realized.

1 <http://public.itrs.net/> – *The link provided at the time the text originally went into print in 2004 has since become inactive. (10 November 2023)*

2 The process described below involves a piece of equipment (CHM) that does not yet exist. The figures presented here are only indicative estimates.

3 E.g. JAVA: <http://java.sun.com>

4 E.g. SGML and its descendants (HTML, XML, etc). Cf.: <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16387>

3. The task to be solved

Digitisation enables convenient access⁵ to cultural documents, facilitating research and education. The drawback of digitised cultural documents is their inherent vulnerability—to the extent that the secretary general of the *European Commission on Preservation and Access* emphasizes the need to maintain the procedures for creating analogue copies alongside digital-only archiving techniques,⁶ even if this limits broader accessibility to the documents.

The *Charter on the Preservation of Digital Heritage*,⁷ adopted by UNESCO on 15 October 2003, calls for the same general problem to be addressed. Article 1 of the *Charter* describes digital heritage as an expanding treasure trove comprising cultural assets that are either digitised versions of previously non-

5 <http://www.knaw.nl/ecpa/members.html#staff> – *The link provided at the time the text originally went into print in 2004 has since become inactive. (10 November 2023)*

6 Yola de Lusenet, *Keeping Things That Work. Preservation Aspects of Digitization*, in Andrea Bozzi, Laura Cignoni, Jean-Louis Lebrave (eds.), *Digital Technology and Philological Disciplines*, Pisa–Roma, Istituti Editoriali e Poligrafici Internazionali, 2004, 145–161.

7 http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html – *The link provided at the time the text originally went into print in 2004 has since become inactive. (10 November 2023)*

digital assets or assets created digitally.⁸ Articles 3 and 4 establish the fact that because of the accelerated obsolescence of machines and programmes and a number of other factors, the loss of the digital heritage will be “rapid and inevitable” unless the Member States find a quick remedy. Article 5 formulates the task for which this invention proposes a solution: “Continuity of the digital heritage is fundamental. To preserve digital heritage, measures will need to be taken throughout the digital information life cycle, from creation to access. Long-term preservation of digital heritage begins with the design of reliable

8 “The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are ‘born digital,’ there is no other format but the digital object.

Digital materials include texts, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral and require purposeful production, maintenance and management to be retained. Many of these resources have lasting value and significance and therefore constitute a heritage that should be protected and preserved for current and future generations. This ever-growing heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression.”

systems and procedures which will produce authentic and stable digital objects.”⁹

The invention aims at the long-lasting preservation of documents representing cultural heritage.

4. The general solution to the problem

4.1 Definitions

4.1.1 *Cultural heritage* is the entirety of humanity’s accumulated cultural wealth, tangible and intangible.¹⁰

4.1.2 The *cultural canon* is a miniature image of cultural heritage; it is cultural wealth as condensed into textbooks.

4.1.3 A **document** is any text. Images, videos, sounds, and other objects converted into text form (e.g. by way of digitisation) are also texts.

4.1.4 A **text** is an ordered set of signs.

4.1.5 A *sign* is what is mapped to an element of a set containing collections of signs.

9 “Continuity of the digital heritage is fundamental. To preserve digital heritage, measures will need to be taken throughout the digital information life cycle, from creation to access. Long-term preservation of digital heritage begins with the design of reliable systems and procedures which will produce authentic and stable digital objects.”

10 http://portal.unesco.org/culture/en/ev.php-URL_ID=2185&URL_DO=DO-TOPIC&URL-SECTION=201.html – *The link provided at the time the text originally went into print in 2004 has since become inactive. (10 November 2023)*

4.1.6 *Exact copying* occurs when the elements of a source text and the elements of the resulting text are mapped consistently to the same elements of a set containing collections of signs.

4.2 Discussion

Within the meaning of the definition of the last term above, only texts can be copied exactly. When copying a text, whatever cannot be copied is not part of the text to begin with. The digitisation of texts is merely a form of exact copying. The digitisation of images and sounds is, then, nothing else but their transformation into text form. Only text databases exist: when incorporating images and sounds into a database, it is not only the metadata on those images and sounds that are entered in the database in a text form, but also the images and the sounds themselves.

In this sense, the age of computers is characterized not so much by the disappearance of texts—the end of the so-called “Gutenberg Galaxy”¹¹—but by the emerging hegemony of text-based communication.

The visible and audible surface is often experienced as a still image, a video, or audio; however, beneath the surface, these too are streams of texts.

11 *Kép, beszéd, írás*, ed. Katalin Neumer, 2003.

Throughout the course of history, the accessibility of texts increases, while the durability of the media carrying those texts decreases.¹²

This general historical rule (which certainly also allows for occasional exceptions) is explained by the fact that ever-broader access to texts is increasingly achieved by means of reproduction (i.e., exact copying in multiple copies).

It was recognised as far back as antiquity that making multiple copies of texts recorded on perishable mediums not only facilitates the spatial distribution of those texts but also contributes to their improved preservation over time—more so than casting them into bronze. As Horace wrote, a masterpiece written on papyrus is “aere perennius,” more lasting than bronze.¹³ He has been proven right: just as an example, over 600 manuscript copies of Cicero’s *De amicitia* and *De officiis* are known today.¹⁴ For the text to survive, the medium only has to survive as long as a copy is made of it. The less energy the copying process requires, the greater the chance that a copy will be made. The same holds true for industrial practice as well: the less energy is required to produce one copy of the textual medium, the easier the

12 Throughout history, there has been a trend of decreasing durability in the medium from petroglyphs and fired clay tablets through papyrus, parchment, the paper used to print incunabula, and newsprint to optical and magnetic devices.

13 Quintus Horatius Flaccus, III, XXX, i.

14 Paolo Fedeli, *Le trappole della contaminazione*, in *Digital Technology and Philological Disciplines*, Pisa–Roma, Istituti Editoriali e Poligrafici Internazionali, 2004, 213.

task of its reproduction. This is why book printing became widespread.

However, the method of preserving texts by reproduction has its drawbacks as well as its advantages. When producing a textual medium by a process of reproduction requires less energy, the medium itself will be more perishable because it will also take less energy to render it unable to retain its text content. Copies produced using a multiplication process of low energy intake will only guarantee the preservation of the text if multiplication does indeed take place regularly. Thus, preserving texts through multiplication presupposes the constant active involvement of society.

This is the explanation behind the historical rule that texts become ever-more broadly accessible and their chances for survival continuously improve despite the diminishing durability of the textual medium. It is thanks to a society that preserves the world of texts through continuous copying that texts surpass the lifetime of their authors and even the languages in which they were written.¹⁵

The Internet does not quite fit the historical rule. Although it has further decreased the energy required for text reproduction, it has not contributed to the multiplication of the number of copies of

¹⁵ Iván Horváth, *On-line Critical Editions in the 90s: Practice and Theory*, in Andrea Bozzi, Laura Cignoni, Jean-Louis Lebrave (eds.), *Digital Technology and Philological Disciplines*, Pisa–Roma, Istituti Editoriali e Poligrafici Internazionali, 2004, 245–258.

the actual text media. In the past, for one to read a book, one had to hold a physical copy of that book in their hand. For a large number of readers to have an essentially simultaneous reading experience of the same book, an equally large number of physical copies of that book had to be available. Today, a single storage device on a single server computer can be accessed by a large number of remote users at essentially the same time. Without going into the details, it can be stated that in the flow of texts over the Internet, a complete copy does not always appear at the receiving end of the text transmission. The copy at the receiving end of the text transmission is selective (hardly any more than the most essential information is ever displayed), fragmentary (it contains no information either on the context or on the overall structure of that context in a wider sense), and it is sometimes not even preserved. Only the copies at the sending end of text transmission may be considered complete, and these are found on the storage devices of the server computer, its mirror servers, and its backups—which means only very few very vulnerable copies exist. Thus, while historically the destruction of physical reproductions of a text has not typically resulted in the loss of the text itself, in the age of the Internet, the destruction of the medium often leads to the loss of the text as well.¹⁶

16 Iván Horváth, *Magyarok Babelben*, Szeged, JATEPress, 2000, 131–135. On the exhibition at the Frankfurt Book Fare of texts that have disappeared from the Internet during the nineties, see <http://magyar-irodalom.elte.hu/contentware/magyar/>

The historical rule implies that texts become ever-more accessible over time. This means that documents representing cultural heritage that are not accessible via the Internet will inevitably lose some of their significance. However, documents that are accessible via the Internet will eventually be lost because while we know of old texts recorded in few copies on durable mediums, and we also know of old texts recorded in many copies on perishable mediums, we do not know of any old texts recorded in few copies on perishable mediums. The Internet cannot guarantee the preservation of digital documents—information created digitally or converted into digital form from existing analogue resources—representing cultural heritage.

While seeking a solution, there are two alternative approaches available: either making the storage devices more durable, or increasing the number of copies.

If the first method is chosen, the vulnerability of the digital documents can be reduced and the dependability of the storage devices can be improved by investing more energy into encoding. This is the approach favoured by large public collections, and it is the main focus of ongoing research. The disadvantage of this method is that it contradicts the historical rule. This is the outdated method employed by ancient Egypt and Persia: painting texts on the walls of hidden tombs deep inside pyramids and carving petroglyphs into rocks on hard-to-reach mountain peaks.

The alternative approach involves multiplying the number of available copies. The advantage of this method is that it aligns with the historical rule. This solution merely complements the existing Internet-connected devices by introducing a novel device—in a few times 10^6 to 10^9 copies—that store an extensive and continuously expanding portion of humanity’s cultural heritage; a device that is centrally written and updated, but that, from the user’s perspective, operates as a read-only device.

4.3 The invention

A storage device written and updated centrally but benefitting end-users as a read-only unit, mass-produced in very large quantities (preferably in millions or billions of copies) and containing digital documents—information created digitally or converted into digital form from existing analogue resources—representing cultural heritage. The provisional name of the device is Cultural Heritage Machine (CHM).

The CHM may be shared among a limited group of users, but is ideally intended for personal use. The following description assumes the single-user scenario.

The CHM is a very large capacity (1 petabyte or more) read/write storage device utilising magnetic, optical, or other technology, and equipped with a standard interface. It does not require power to retain the data recorded for at least several years. Characterised by low energy consumption and a low

noise level, it comes in a compact form factor and, if mass-produced, at a relatively low price point. While it can be equipped with its own operating platform and a smart user interface,¹⁷ its basic use case is to be connected to the user's own personal networked communication device(s) through a standard interface. It can be marketed either as integrated into other devices or as a stand-alone module. It does not have to fulfill any special requirements of in terms of reliability or durability.

The data stored on the CHM is relatively structured, and can be organized and searched using the same type of user interface that would be employed for accessing the same data on the network. The content of the CHM can use network hyperlinks without restrictions.

The CHM storage device consists of a core segment and a supplementary segment. The core segment is not user-writable. Data can be written into the CHM devices by an institution authorised by society to exercise such right. The data are also updated by these institutions or by their representatives. It is desirable that such updates do not merely overwrite earlier data, but that they preserve the historical sequence of such updates; all consecutive data update layers should be recoverable and consultable.

The supplementary (non-core) segment of the CHM storage device is only writable for the user. This supplementary segment can be physically lo-

¹⁷ This might especially be useful in the case of old database programs that run in an emulated environment.

cated outside the CHM. Its user interface is the same as that of the network and that of the core segment of the CHM.

The documents stored in the core segment of the CHM device can be divided into three groups: documents forming part of the national cultural heritage, documents forming part of the international cultural heritage, and documents forming part of general cultural heritage. The institution endowed with the right to curate the data content of the CHM devices applies moderate filtering criteria in the case of the first group of documents and stricter selection criteria in the case of the second group. Documents that feature in the first group for a specific nation may be made available to other nations as part of their second group of documents, ensuring that the contents of the personal CHM devices used by members of different nations are complementary to one another. In the case of documents forming part of general cultural heritage that cannot be assigned to any specific nation, moderate filtering criteria must be applied.

Documents representing cultural heritage can be digitised for the purposes of the CHM at various resolutions: sensorially acceptable, sensorially authentic, or far beyond sensorial authenticity. It is desirable that resolution improve at the same rate as the storage capacity of the CHM devices increases over time.

The CHM devices apply data storage and data addressing methods that adhere to international

standards, enabling CHM users to exchange their observations concerning the cultural heritage in a mutually intelligible manner.

4.4 The CHM as a teaching resource

The CHM device—described above in 4.3 as the invention—doubles as a teaching resource.

What we call the cultural canon is only a small fragment of the entirety of humanity’s cultural heritage—that which is incorporated into our textbooks. While the scope of the cultural heritage is continuously expanding, the scope of the canon is decreasing.

Some proponents in the field of educational research advocate for a significant further reduction of the cultural canon. Instead, they suggest shifting the focus towards developing the problem-solving skills that students require for practical applications in their lives. Others maintain that the traditional cultural curriculum does indeed have an important role to play in elite education and should therefore be kept.¹⁸ The CHM device can be seamlessly integrated into both approaches as a teaching resource.

As a matter of course, students receiving elite education possess—and even as adults, will retain—the key to engaging with and appreciating the cultural heritage made available to them through the CHM,

18 *Tartalmak és módszerek az ezredforduló iskolájában. Tanulmányok a tantárgyi helyzetfelmérésről*, ed. Zoltán Kerber, 2004, especially 45–67.

which serves as an infinite resource for their continuous self-education. In turn, students outside the system of elite education—and later on as adults—who may not have the means or the desire to cultivate themselves independently by absorbing and internalising the cultural heritage can still possess a personal copy of it in the physical form of the CHM, an object they have the skills to use.

In both cases, the main task of the school system is to familiarise students with the operating procedures of the CHM, offering them opportunities to acquire experience in engaging with the cultural heritage. In all cases, the CHM requires active participation from the student and an entirely new pedagogical methodology from the teacher. The objective is to provide every individual, as early as in their childhood, possibly in a ritualized form, with a personal CHM device that they will keep throughout their lives as a constant personal companion, much like their eyeglasses or cell phones.

Note. Both legislation and the business sector can play important roles in promoting the widest possible use of the CHM device, each contributing in their own distinct ways. Various approaches are already available for the copyright clearance of the documents stored on CHM devices.¹⁹

19 Iván Horváth, *Művelődés közpénzen*, in HVG, 9 June 2001. Bálint Magyar, *Network of Clearinghouses for Educational Software*, Rome, January 2004 (presentation): <http://magyar-irodalom.elte.hu/biop/barbar/NCES.PDF>

In accordance with 4.1.5 and 4.1.6 above, documents can only be understood if the individual is in possession of the relevant set of collections of signs. The preservation of the cultural heritage only makes sense as long as there are human beings who have an avid interest in engaging with it and a willingness to receive this heritage. Consequently, the two tasks described under 4.3 and 4.4 above (the preservation of the cultural heritage and its transmission in the educational context with the help of the CHM device) are interdependent and presuppose one another.

As a massive potential market, the educational system would make the mass-production of the CHM economically feasible. In turn, the mass-production of the CHM would facilitate the preservation of digital documents—information created digitally or converted into digital form from existing analogue resources—representing cultural heritage.

5. Patent claims

5.1 A process for the long-term preservation of digital documents–information created digitally or converted into digital form from existing analogue resources–representing cultural heritage, such process being characterized by increasing the quantity, not the dependability, of the storage devices.

5.2 A process for the application of the method described in 5.1 above for the pedagogical transmission of digital documents–information created digitally or converted into digital form from existing analogue resources–representing cultural heritage, such process being characterized by the use of the storage devices therein described as teaching resources.

Címlap és tipográfia
STÁHL ESZTER
Szerkesztette
HERNER JÁNOS

A szöveg korábban nyomtatott formában is megjelent
az alábbi kiadványban:

Horváth Iván: *Gépeskönyv*. Budapest, 2006.

Previously published in hard copy in:

Iván Horváth: *Gépeskönyv*. Budapest, 2006.

Ez a kiadvány ingyenesen hozzáférhető minden fontosabb
digitális könyvtárban és repozitóriumban.

This publication is available free of charge to all in all major
digital libraries and repositories.

A kiadásért a Q.E.D. Kiadó ügyvezetője felel.

All responsibility for this publication is assumed by the
Managing Director of the Publishing House Q.E.D.

© Q.E.D. Kiadó, 2023