

JAKUB SWACHA

POPRAWA EFEKTYWNOŚCI PRZECHOWYWANIA DANYCH W SYSTEMIE ARCHIWIZACJI PRAC DYPLOMOWYCH*

1. Przedmiot badań

Studenci kierunków informatycznych Wydziału Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego zobowiązani są złożyć wykonane przez siebie prace dyplomowe (licencjackie i magisterskie) w dwóch postaciach – wydrukowanej oraz elektronicznej. Te ostatnie są przechowywane w archiwum Instytutu Informatyki w Zarządzaniu.

Najważniejszym przedmiotem archiwizacji jest plik zawierający treść pracy dyplomowej, przy czym zaleca się, by format tego pliku był zgodny z Microsoft Word 6.0. Wychodząc naprzeciw oczekiwaniom studentów, akceptowane są także pliki zapisane w formatach innych wersji programu Word, w kompatybilnym z programem Word formacie Microsoft Rich Text Format, a nawet innych popularnych edytorów tekstu. Z możliwości tej korzysta niewielka grupa studentów (tabela 1).

* Praca naukowa finansowana ze środków na naukę w latach 2007–2008 jako projekt badawczy.

Tabela 1

Formaty pliku z tekstem pracy dyplomowej

Format	Oznaczenie	Odsetek prac
Microsoft Word 6.0 lub kompatybilny	.doc	86
Portable Document Format	.pdf	6
Microsoft Rich Text Format	.rtf	4
Microsoft Word 2007	.docx	2
OpenOffice/StarOffice	.odt	2

Źródło: badania własne.

Przedstawione w tabeli 1 wyniki dotyczą pięćdziesięciu losowo wybranych prac złożonych w latach 2001–2008. Mając na uwadze niewielki okres, jaki upłynął od premiery Microsoft Word 2007, należy się spodziewać, że niewielki obecnie udział tego formatu w liczbie prac składanych przez studentów ulegnie w przyszłości zwiększeniu.

Prócz pliku zawierającego treść pracy dyplomowej archiwizacji podlegają załączniki elektroniczne. Prace czysto teoretyczne zwykle ich nie mają, w pracach stosowanych, w zależności od tematyki, mogą one zawierać:

- dodatkową dokumentację rezultatów uzyskanych w pracy i sposobów ich uzyskania (tabele, schematy, zdjęcia, opisy),
- oprogramowanie (i jego kod źródłowy) wykonane przez autora pracy,
- dodatkowe oprogramowanie, zwykle nienapisane przez autora pracy, konieczne do uruchomienia oprogramowania autorskiego,
- dokumentację oprogramowania wykonanego przez autora pracy,
- bazę danych,
- specyficzne dla tematu pracy dane, wykorzystane jako dane wejściowe lub stanowiące jej rezultat (dane wyjściowe),
- oprogramowanie konieczne do odczytania (lub przetwarzania) załączonych danych, zwykle nienapisane przez autora pracy,
- instrukcję obsługi oprogramowania,
- prezentację multimedialną,
- bazę bibliograficzną dokumentującą źródła wykorzystane w pracy,
- prezentację multimedialną treści pracy (przeznaczoną do przedstawienia podczas obrony).

Obecnie nie jest egzekwowany wymóg dołączenia załączników, stąd wśród badanych w wypadku zaledwie 8% zarchiwizowanych zostało wraz z jakimkolwiek załącznikiem, a tylko 4% z załącznikami stanowiącymi wyczerpujące uzupełnienie treści pracy.

Prace poddawane archiwizacji w Instytucie Informatyki w Zarządzaniu na Wydziale Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego są pracami studentów następujących rodzajów studiów:

- a) na kierunku informatyka i ekonometria:
 - studia stacjonarne magisterskie,
 - studia niestacjonarne licencjackie,
 - studia niestacjonarne magisterskie,
- b) studia podyplomowe na różnych kierunkach informatycznych.

W archiwum znajdują się prace studentów odbywających studia w Szczecinie i w zamiejscowych ośrodkach dydaktycznych. Począwszy od roku akademickiego 2007/2008 w przypadku studiów stacjonarnych i 2008/2009 w przypadku studiów niestacjonarnych powołano studia inżynierskie w specjalności inżynieria Internetu. Ponadto wydzielono pierwszy stopień studiów stacjonarnych licencjackich na kierunku informatyka i ekonometria. Biorąc pod uwagę wprowadzone zmiany, przyjęto następujące założenia co do liczby prac archiwizowanych w kolejnych latach:

- 2009 – 250,
- 2010 – 300,
- 2011 i później – 400.

Objętość pamięci potrzebna do przechowania pojedynczej pracy zależy od wielkości pliku zawierającego treść pracy dyplomowej oraz wielkości załączników. Jest ona bardzo zróżnicowana i zależy nie tylko od liczby stron w pracy, ale przede wszystkim od liczby ilustracji i ich typu (zdjęcia, zrzuty ekranu – grafika rastrowa; szkice, diagramy – grafika wektorowa). Wśród badanych prac najmniejsza miała 0,6 megabajta długości, podczas gdy największa – 157,8 megabajta. Średnia wielkość pliku zawierającego treść pracy dyplomowej wynosiła 7,1 megabajta. Założywszy wykładniczy wzrost wielkości pracy, przy użyciu narzędzi analizy regresji programu Microsoft Excel, oszacowano średni roczny wzrost wielkości pojedynczej pracy w latach 2001–2008 na poziomie 16,8% rocznie. Ze względu na bardzo duże różnice w długości prac w jednym roku, trudno jednoznacznie stwierdzić, na ile liczba ta ilustruje faktyczną tendencję, a na ile wynika ze specyfiki prac bronionych w kolejnych latach (zmieniająca się

popularność różnych tematów, formaty graficzne, w których zakodowano osadzone w pracy ilustracje). Istnieje przynajmniej jedna przesłanka wskazująca na to, że przy tej samej liczbie stron w pracy, długość pliku zawierającego treść pracy dyplomowej powinna rosnać: mianowicie, jest nią coraz większa rozdzielczość umieszczanych w pracy ilustracji. Obrazy o dużej rozdzielczości zajmują więcej pamięci niż o niskiej, ponadto duża rozdzielczość pozwala zamieścić obrazy, które przy niskiej musiałyby być pominięte ze względu na ich małą czytelność. Stąd, mimo wymienionych wątpliwości, w szacunkach rocznego wzrostu wielkości pojedynczej pracy przyjmujemy obliczony współczynnik 16,8%.

Znaczenie załączników jest z pewnością większe w wypadku prac inżynierskich, w związku z czym w przyszłości należy się spodziewać bardziej ścisłego egzekwowania wymogu składania prac wraz z załącznikami. Oczywiście, nie każdy temat pracy wymaga załącznika. Analizując tematykę badanych prac (wśród których nie było prac inżynierskich), a także biorąc pod uwagę liczbę prac inżynierskich zawierających załączniki elektroniczne wśród prac bronionych w Wyższej Informatycznej Szkole Zawodowej w Gorzowie Wielkopolskim, przyjęto, że przy ścisłych wymogach dotyczących składania załączników, odsetek prac je zawierających należy docelowo szacować (od roku 2011) na 75%, a w latach 2009–2010 odpowiednio na 25 i 50%. Za przewidywaną średnią wielkość załączników (w pracach, które je mają) przyjęto średnią wielkość załączników wśród prac, w których stanowiły one wyczerpujące uzupełnienie treści pracy. Wynosiła ona 177,2 megabajta.

Z braku możliwości uzyskania rzetelnych szacunków (zbyt mała liczba prac złożonych wraz z pełnymi załącznikami), przyjęto współczynnik wzrostu wielkości załączników identyczny ze współczynnikiem wzrostu wielkości samej pracy, to jest 16,8%.

Na podstawie tych szacunków otrzymano łączną prognozę przyrostu objętości danych w archiwum, co przedstawiono w tabeli 2. Całkowitą objętość archiwum otrzymano, przyjmując, że na początku rozpatrywanego okresu zawierało ono tysiąc prac o łącznej objętości 19 gigabajtów.

Tabela 2

Przewidywany przyrost danych przechowywanych w archiwum

Rok	Liczba prac	Przyrost objętości	Łączna pojemność
2009	1250	12,9	31,9
2010	1550	33,5	65,4
2011	1950	76,4	141,8
2012	2350	89,2	231,0

Źródło: badania własne.

Obecnie prace przechowywane są na nośnikach dostarczonych przez studentów, najczęściej na dyskach optycznych CD-R i DVD-R. Z jednej strony zwalnia to instytut od ponoszenia kosztów zakupu dodatkowych nośników danych, z drugiej ma wiele wad, jak na przykład:

- brak kopii zapasowych: uszkodzenie oryginalnego nośnika (nierzadko niskiej jakości) oznacza utratę danych,
- konieczność ręcznego wyszukiwania nośnika dla podanego autora lub tytułu pracy,
- brak pełnej wiedzy o załącznikach dołączonych do zebranych prac,
- brak możliwości wyszukiwania pełnotekstowego w tekście prac i załącznikach tekstowych (na przykład w kodzie źródłowym) zgromadzonych w archiwum,
- brak możliwości prowadzenia analizy statystycznej całości zasobów zgromadzonych w archiwum.

Prace bronione w latach 2002–2005 były wprowadzane do systemu bazodanowego (wykorzystującego jako podstawowy nośnik danych dyski twarde) przez system Rodan OfficeObjects. Do bazy dołączano plik z treścią pracy, załączniki (w omawianym okresie miały one zwykle niewielką objętość, o czym świadczy choćby fakt, że większość prac – wraz z załącznikami – składano wtedy na dyskietkach o pojemności 1,44 megabajta) oraz krótką informację pozwalającą na odszukanie prac w bazie (tytuł, autor, promotor, tematyka). Opinie pracowników były pozytywne, jeśli chodzi o sposób archiwizowania prac¹, możliwość ich przeglądania była przydatna w trakcie prowadzonych seminariów dyplomowych (możliwość łatwego odniesienia przedstawianych przez studentów treści do prac wcześniej bronionych w instytucie) tworzyła okazję do interesujących badań

¹ Na podstawie badań własnych.

naukowych, dla których bazę analityczną stanowiło archiwum prac dyplomowych².

Splot niekorzystnych okoliczności – odejście z pracy osób zajmujących się administracją archiwum przy dużym obłożeniu obowiązkami pozostałych pracowników administracyjno-biurowych; problemy sprzętowe; nawiązanie przez wydział współpracy z firmą plagiat.pl, które podważyło jedno z ówczesnych założeń funkcjonowania archiwum dyskowego, jakim było wyszukiwanie plagiatów – sprawił, że w roku 2006 zaprzestano umieszczania w systemie kolejnych prac i przechowywano je jedynie na dostarczonych przez studentów nośnikach. To utrudnia korzystanie z archiwum pracownikom instytutu.

Celem niniejszego artykułu jest:

- wykazanie, że archiwizacja prac na nośnikach o szybkim dostępie ma nie tylko walory funkcjonalne, ale poprawia także efektywność ekonomiczną przechowywania danych, zmniejszając koszty dostępu do nich przez zmniejszenie czasu traconego na ich wyszukiwanie i odczytywanie;
- wykazanie, że zastosowanie kompresji danych może znacząco ograniczyć koszty użytkowania archiwum;
- znalezienie najbardziej efektywnego ekonomicznie wariantu budowy systemu archiwizacji prac dyplomowych w perspektywie najbliższych czterech lat.

Do osiągnięcia tych celów wykorzystana zostanie metodyka oceny efektywności ekonomicznej przechowywania danych³. Przedstawione tu wyniki stanowią zatem interesujący przyczynek do oceny stosowalności metodyki, biorąc pod uwagę specyfikę badanego systemu przechowywania danych, a przede wszystkim fakt, że użytkowany jest on przez organizację niedziałającą w celu osiągnięcia zysku.

² Zob. np. M. Mikołajczyk, J. Swacha, *Computer-Aided Paper Evaluation*, “Image Analysis, Computer Graphics, Security Systems and Artificial Intelligence Applications” 2005, vol. II, Białystok.

³ J. Swacha, *Zarządzanie przechowywaniem danych. Metodyka oceny efektywności*, Wydawnictwo Placet, Warszawa 2009.

2. Warunki ekonomiczne funkcjonowania archiwum

Efektywność to relacja między całkowitymi nakładami a efektami⁴. W przypadku systemu przechowywania danych jej miarą może być na przykład stosunek całkowitych nakładów poniesionych na jego budowę, modernizację i utrzymanie do całkowitych efektów jego działania.

Można wyróżnić wiele rodzajów nakładów związanych z przechowywaniem danych⁵. Tu ograniczymy ich klasyfikację do następujących grup:

- koszty nośników danych oraz sprzętu i oprogramowania niezbędnego do odczytu i zapisu danych, w szczególności:
 - o koszty zakupu (cena zakupu, koszt dostawy, koszt instalacji, koszt niezbędnych szkoleń i/lub wsparcia technicznego),
 - o koszty utrzymania (koszty serwisu i napraw, energii elektrycznej, utrzymanie powierzchni użytkowej pomieszczeń fizycznie zajętej przez urządzenia – czynsz, ochrona, sprząatanie, oświetlenie, centralna klimatyzacja itp.);
- koszty obsługi (wynagrodzenie pracowników odpowiedzialnych za administrację sprzętem i oprogramowaniem oraz wykonujących ich bezpośrednią obsługę, jeżeli taka jest konieczna);
- koszty dostępu do danych (koszt czasu dla użytkowników korzystających z systemu);
- koszty niedostępności danych (czasowej lub całkowitej utraty danych – ich odtworzenia, zastąpienia lub odstąpienia od realizacji czynności, do których były niezbędne).

W przypadku archiwum prac dyplomowych:

- koszty nośników danych, na których są składane prace przez studentów, nie obciążają budżetu jednostki, a zatem mogą być pominięte;
- koszty wtórnych nośników danych (na które prace są przenoszone po ich złożeniu) oraz sprzętu (prosty sieciowy serwer plików oparty na komputerze klasy PC) i oprogramowania niezbędnego do odczytu i zapisu danych (wyłącznie oprogramowanie systemowe) zostaną przyjęte według cennika firmy

⁴ J. Kisielnicki, H. Sroka, *Systemy informacyjne biznesu*, Wydawnictwo Placet, Warszawa 2005, s. 325.

⁵ B. Rogers, *Calculating the ROI of Information Lifecycle Management*, Storage Networking Industry Association 2006, <http://www.snwusa.com/documents/presentations-f06/BobRogers.pdf>, s. 22.

- Flop (www.flop.pl), z którą Uniwersytet Szczeciński zawarł umowę na dostawę sprzętu komputerowego;
- wśród kosztów utrzymania wzięto pod uwagę jedynie koszty energii elektrycznej, przyjmując zużycie według specyfikacji producentów;
 - koszty obsługi, w przypadku wykorzystania nośników wtórnych:
 - administracja serwerem sieciowym, wprowadzanie aktualizacji oprogramowania systemowego, usuwanie ewentualnych usterek to koszt zatrudnienia pracownika technicznego, jeśli chodzi o czas pracy – dwie godziny w miesiącu, biorąc pod uwagę średnie wynagrodzenie brutto pracowników technicznych, miesięczny koszt tych działań oszacowano na 16 złotych miesięcznie;
 - wprowadzanie na nośniki wtórne prac z dostarczonych przez studentów nośników wymiennych – realizowane przez pracowników administracyjno-biurowych – czas pracy: około trzy minuty, koszt: 50 groszy za każdą taką pracę;
 - koszt dostępu do danych:
 - w odniesieniu do prac przechowywanych na dostarczonych przez studentów nośnikach wymiennych każdorazowy dostęp wymaga procedury analogicznej, jak w przypadku dostępu do wydruku pracy (patrz niżej), stąd koszty oszacowano na 5 złotych;
 - w odniesieniu do prac przechowywanych na nośnikach wtórnych, koszt dostępu można uzyskać, mnożąc czas dostępu do przeciętnej wielkości pracy przez średnie wynagrodzenie brutto pracowników naukowo-dydaktycznych; oszacowano na 20 złotych na godzinę;
 - koszty braku dostępu do danych:
 - czasowy – z uwagi na to, że dostęp do zarchiwizowanych prac nie ma znaczenia krytycznego dla instytucji, a przewidywany czas przestoju serwera plików wynosi mniej niż jedną godzinę na rok, koszty te pominięto w dalszych rachunkach;
 - trwałej utraty danych – biorąc pod uwagę liczbę prac, koszt ponownego wprowadzenia prac z nośników wymiennych na nośniki wtórne i prawdopodobieństwo awarii dysku, wynoszące mniej niż 10^{-5} , wartość oczekiwana strat z tego tytułu wynosi mniej niż 1 złoty rocznie; dlatego również te koszty pominięto w dalszych rachunkach; należy zwrócić uwagę, że zachowanie 100% prac w omawianym archiwum nie jest konieczne, co odróżnia je od wielu innych archiwów.

Ponadto, przyjęto:

- roczny wzrost cen energii elektrycznej w wysokości 25%,
- roczny wzrost wynagrodzeń brutto w wysokości 5%.

Korzyści z przechowywania prac dyplomowych mogą być wielorakie, tu rozpatrzmy dwie podstawowe sytuacje:

- gdy dostęp do elektronicznej wersji pracy może być zastąpiony dostępem do jej wydruku (na przykład, gdy jego celem było zapoznanie się z zawartością pracy);
- gdy dostęp do elektronicznej wersji pracy nie może być zastąpiony dostępem do jej wydruku (na przykład, gdy jego celem była analiza statystyczna tekstu pracy lub zapoznanie się z zawartością jej elektronicznych załączników).

Z uwagi na to, że działalność organizacji nie jest nastawiona na osiąganie zysku, za wartość korzyści w pierwszym przypadku przyjmiemy wartość kosztów związanych z dostępem do wydruku pracy, które musiałyby być poniesione, gdyby nie była dostępna wersja elektroniczna pracy.

Wydruki prac przechowywane są w zamkniętym archiwum, do którego dostęp możliwy jest jedynie wraz z pracownikiem administracyjno-biurowym. Z uwagi na sposób oprawiania wydruków prac przez studentów (okładki nieposiadające na grzbiecie informacji o autorze i tytule pracy), mimo że prace przechowywane w archiwum są uporządkowane, wyszukanie pojedynczej pracy zajmuje około dziesięciu minut. W tym czasie nie jest możliwe wykonywanie innych czynności ani przez pracownika administracyjno-biurowego (który wyszukuje prace), ani przez pracownika naukowo-dydaktycznego (który zlecił ich wyszukanie). Uwzględniając średnie wynagrodzenie brutto obu tych grup pracowników w przeliczeniu na godzinę pracy, przyjęto, że koszt dostępu do wydruku pracy wynosi 5 złotych. Z uwagi na to, że w drugim przypadku (gdy dostęp do elektronicznej wersji pracy nie może być zastąpiony dostępem do jej wydruku) nie ma możliwości wykonania podobnych obliczeń, przyjęto w nim identyczny koszt.

W konsekwencji, zakładając, że pracownicy naukowo-dydaktyczni korzystają z dostępu do wydruków prac, wartość korzyści z takiego dostępu nie jest mniejsza od ponoszonych kosztów, a wartość korzyści z pojedynczego dostępu do pracy określono na 5 złotych.

3. Wskaźniki ekonomiczne funkcjonowania archiwum

W celu obliczenia wskaźników przyjęto następujące założenia eksploatacyjne:

- okres eksploatacji sprzętu – 4 lata,
- liczba operacji dostępu do prac zgromadzonych w archiwum w ciągu roku:
 - o w celu zapoznania się z treścią prac – 500,
 - o w celu zebrania danych do badań – 5 na każdą pracę w archiwum.

Wybrano następującą konfigurację serwera plików: dwurdzeniowy procesor Intel Core 2 Duo 6420 2,13 GHz (FSB 1066 MHz), płyta główna oparta na układzie Intel P965 / ICH8, pamięć DDR2 800 MHz 2 GB, pod kontrolą 32-bitowego systemu operacyjnego Windows XP Professional, koszt zakupu w roku 2008 – 2000 złotych.

Za pierwotny nośnik danych przyjęto DVD-R. Rozważono następujące rodzaje wtórnych nośników danych:

- zewnętrzny dysk twardy Western Digital Passport 120 GB (5400 obrotów na minutę), oznaczany dalej symbolem WDP120 (koszt zakupu 220 złotych),
- wewnętrzny dysk twardy Seagate Barracuda 320 GB (7200 obrotów na minutę), oznaczany dalej symbolem SEA320 (koszt zakupu 250 złotych),
- wewnętrzny dysk twardy Western Digital Raptor 150 GB (10 000 obrotów na minutę), oznaczany dalej symbolem WDR150 (koszt zakupu 575 złotych).

Przyjęto, że dyski kupowane będą parzyście, przy czym dyski wewnętrzne łączone będą w trybie RAID 0, natomiast dyski zewnętrzne działać będą jako JBOD⁶.

Rozważono przechowywanie prac bez kompresji oraz kompresję przy użyciu trzech metod: Deflate (program kompresujący Info-Zip 2.3, opcja -5), LZMA i PPMII (program kompresujący: 7-Zip 4.57, opcje -mx3 i -m0=PPMd: o5). Pomiar efektywności kompresji oraz czasu dekompresji dla plików zawierających treść pracy wykonano dla 50 losowo wybranych prac dyplomowych. Pomiar efektywności kompresji oraz czasu dekompresji dla plików zawierających załączniki do pracy, z uwagi na zbyt małą liczbę zgromadzonych dotąd załączników, wykonano dla 10 plików referencyjnych z zestawu Maximum

⁶ J. Swacha, *Zarządzanie...*, s. 96.

Compression⁷, które reprezentują typy danych często spotykane w załącznikach do prac. W tabeli 3 przedstawiono zestawienie uzyskanych stopni kompresji i czasów odczytu (zawierające czasy dekompresji). Wyniki uzyskano przy użyciu programu CoTe⁸.

Tabela 3

Uzyskane średnie stopnie kompresji i czasy odczytu

Metoda kompresji	Stopień kompresji		Pamięć masowa	Czas odczytu	
	treść pracy	załączniki		treść pracy	załączniki
brak	1		WDP120	39,0	
			SEA320	33,8	
			WDR150	33,3	
Deflate	3,01	3,89	WDP120	13,0	13,8
			SEA320	11,6	12,3
			WDR150	11,5	12,2
LZMA	4,05	4,90	WDP120	46,7	40,9
			SEA320	45,6	39,8
			WDR150	45,5	39,7
PPMII	3,37	6,21	WDP120	378,7	268,6
			SEA320	377,8	267,7
			WDR150	377,7	267,6

Źródło: badania własne. Czas odczytu 1 GB danych podano w sekundach.

Przy założonej liczbie i wielkości prac dane z tabeli 3 przekładają się na wzrost objętość archiwum, tak jak to podano w tabeli 4.

Tabela 4

Przyrost objętości danych z uwzględnieniem kompresji

Metoda kompresji	2009	2010	2011	2012
Brak	31,9	65,4	141,8	231,0
Deflate	17,2	45,5	97,2	136,8
LZMA	13,5	35,9	76,8	108,1
PPMII	11,6	29,6	62,5	88,0

Źródło: badania własne.

⁷ W. Bergmans, *Maximum Compression: Lossless data compression software benchmarks/comparisons*, <http://www.maximumcompression.com>, 17 May 2008.

⁸ J. Swacha, *Benchmarking data compression and storage performance with the CoTe tool*, „Metody Informatyki Stosowanej” 2008, no. 14, s. 269–276.

Jak widać, przy zakupie dwóch dysków uzyskana pojemność będzie wystarczająca do zaspokojenia co najmniej czteroletnich potrzeb archiwum (tabela 5).

Tabela 5

Zaspokojenie potrzeb pamięciowych w przyszłości

Pamięć masowa	Metoda kompresji	Wykorzystanie pojemności w 2012 [%]	Rok, w którym konieczne będzie powiększenie zasobów
WDP120	Brak	96	2012
	Deflate	57	2013
	LZMA	45	2014
	PPMII	37	2014
SEA320	Brak	36	2014
	Deflate	21	2014
	LZMA	17	2015
	PPMII	14	2015
WDR150	Brak	77	2012
	Deflate	46	2014
	LZMA	36	2014
	PPMII	29	2014

Źródło: badania własne.

Uwzględniając czasy odczytu podane w tabeli 3 i przyjęte założenia, koszty dostępu do danych przedstawiają się tak jak to zaprezentowano w tabeli 6.

Należy mieć na uwadze, że celem porównania, wartość kosztów wyliczona dla przypadku, w którym prace nie są przenoszone na nośniki wtórne, uwzględnia identyczną liczbęostępów, jak w sytuacji, gdy są one przenoszone na dyski twarde. W praktyce żmudność dostępu do prac zapisanych na nośnikach wymiennych jest na tyle duża, że faktyczna liczbaostępów jest dużo mniejsza. W tabeli 7 zawarto koszty łączne utworzenia i użytkowania archiwum.

Tabela 6

Koszty dostępu do danych [zł]

Pamięć masowa	Metoda kompresji	2009	2010	2011	2012	Suma
nośniki wymienne	nieistotna	33750	41250	51250	61250	187500
WDP120	Brak	75	210	468	685	1438
	Deflate	25	70	156	228	479
	LZMA	90	251	560	821	1722
	PPMII	730	2037	4541	6656	13963
SEA320	Brak	65	182	405	594	1246
	Deflate	22	62	139	204	428
	LZMA	88	245	547	801	1681
	PPMII	728	2032	4530	6640	13930
WDR150	Brak	64	179	399	585	1228
	Deflate	22	62	138	202	424
	LZMA	88	245	546	800	1678
	PPMII	728	2032	4529	6638	13927

Źródło: badania własne.

Tabela 7

Całkowite koszty utrzymania archiwum [zł]

Pamięć masowa	Metoda kompresji	2009	2010	2011	2012	Suma
nośniki wymienne	nieistotna	33 750	41 250	51 250	61 250	187 500
WDP120	brak	618	576	909	1 151	3 255
	Deflate	568	437	597	694	2 296
	LZMA	633	618	1 002	1 286	3 539
	PPMII	1 273	2 404	4 982	7 121	15 780
SEA320	brak	649	562	864	1 081	3 156
	Deflate	606	443	598	691	2 338
	LZMA	672	626	1 006	1 288	3 591
	PPMII	1 312	2 413	4 989	7 127	15 840
WDR150	brak	972	558	856	1 070	3 457
	Deflate	930	441	595	687	2 653
	LZMA	996	624	1 003	1 285	3 907
	PPMII	1 636	2 411	4 986	7 123	16 156

Źródło: badania własne.

Jak widać, w perspektywie czteroletniej najtańsze rozwiązanie oparte jest na dyskach zewnętrznych WD Passport, charakteryzujących się niskim poborem energii elektrycznej, jak również niską ceną jednostkową; zbliżone koszty łączne uzyskano dla rozwiązania opartego na dyskach wewnętrznych Seagate Barracuda.

Dla łatwiejszego porównania poszczególnych wariantów, w tabeli 8 zawarto współczynniki efektywności obliczone jako stosunek korzyści do nakładów.

Tabela 8

Współczynniki efektywności przechowywania danych

Pamięć masowa	Metoda kompresji	2009	2010	2011	2012
WDP120	brak	54,6	71,6	56,4	53,2
	Deflate	59,4	94,5	85,8	88,3
	LZMA	53,3	66,8	51,2	47,6
	PPMII	26,5	17,2	10,3	8,6
SEA320	brak	52,0	73,4	59,3	56,7
	Deflate	55,7	93,2	85,7	88,7
	LZMA	50,2	65,9	51,0	47,5
	PPMII	25,7	17,1	10,3	8,6
WDR150	brak	34,7	73,9	59,8	57,2
	Deflate	36,3	93,5	86,1	89,1
	LZMA	33,9	66,1	51,1	47,7
	PPMII	20,6	17,1	10,3	8,6

Źródło: badania własne.

Podsumowanie

Uzyskane wyniki wskazują, że

- przechowywanie danych (w tym przypadku prac dyplomowych) na pamięciach masowych niewymagających ręcznej obsługi jest uzasadnione ekonomicznie, jeżeli weźmie się pod uwagę koszty obsługi, jakie należałoby ponieść dla zapewnienia tego samego poziomu wykorzystania przechowywanych danych;
- w przypadku przechowywania danych o umiarkowanej liczbieostępów, dużo większy wpływ na koszty użytkowania archiwum od wydajności pamięci masowej ma wykorzystana metoda kompresji danych; metody zapew-

niające szybką dekompresję danych wykazują znacznie wyższą przydatność niż metody charakteryzujące się lepszymi współczynnikami kompresji;

- odpowiednio dobierając typ pamięci masowej i metodę kompresji, można wielokrotnie zredukować koszty przechowywania danych.

Przeprowadzone badania pokazały także, że możliwe jest wykorzystanie metodyki oceny ekonomicznej efektywności przechowywania danych także do oceny systemów przechowywania danych instytucji niekomercyjnych, pod warunkiem przyjęcia właściwych założeń co do wartości korzyści wynikających dla tych instytucji z tego tytułu.

Literatura

- Bergmans W., *Maximum Compression: Lossless data compression software benchmarks/comparisons*, <http://www.maximumcompression.com>, 17 May 2008.
- Kisielnicki J., Sroka H., *Systemy informacyjne biznesu*, Wydawnictwo Placet, Warszawa 2005.
- Mikołajczyk M., Swacha J., *Computer-Aided Paper Evaluation*, "Image Analysis, Computer Graphics, Security Systems and Artificial Intelligence Applications" 2005, vol. II, Białystok.
- Rogers B., *Calculating the ROI of Information Lifecycle Management*, Storage Networking Industry Association, 2006, <http://www.snwusa.com/documents/presentations-f06/BobRogers.pdf>.
- Swacha J., *Benchmarking data compression and storage performance with the CoTe tool*. „Metody Informatyki Stosowanej” 2008, no. 14.
- Swacha J., *Zarządzanie przechowywaniem danych. Metodyka oceny efektywności*, Wydawnictwo Placet, Warszawa 2009.

IMPROVING THE DATA STORAGE EFFICIENCY IN A DIPLOMA THESES ARCHIVIZATION SYSTEM

Summary

Since many years ago the Institute of Information Technology in Management in the Faculty of Economic Sciences and Management of the University of Szczecin stores both the text and attachments of submitted B.Sc. and M.Sc. theses in an electronic

form. In this paper an attempt has been made to point at the data storage technological solutions that will be capable, in the future, of limiting the costs of these archiving, at the same time allowing fast access to their content. For this purpose the author's methodology for assessment of data storage economic efficiency is used, which also allows to show its usability in case of being applied at a non-commercial educational institution.

Translated by Jakub Swacha