

**ДЕРЖАВНИЙ КОМІТЕТ АРХІВІВ УКРАЇНИ  
УКРАЇНСЬКИЙ НАУКОВО-ДОСЛІДНИЙ ІНСТИТУТ АРХІВНОЇ  
СПРАВИ ТА ДОКУМЕНТОЗНАВСТВА  
ЦЕНТРАЛЬНИЙ ДЕРЖАВНИЙ ЕЛЕКТРОННИЙ АРХІВ УКРАЇНИ**

**КРИТЕРІЇ ВІДБОРУ ВЕБ-РЕСУРСІВ  
ЯК СКЛАДОВОЇ НАФ І ДЖЕРЕЛА КОМПЛЕКТУВАННЯ  
ЦЕНТРАЛЬНОГО ДЕРЖАВНОГО  
ЕЛЕКТРОННОГО АРХІВУ УКРАЇНИ**

**Методичні рекомендації**

**Схвалено  
Нормативно-методичною комісією  
Держкомархіву України**

**Протокол № 4 від 28.10.2009 р.**

**Укладачі: д.і.н., зав. відділом С.Г. Кулешов  
к.т.н., с. н. с. П.М. Марченко**

**Київ – 2009**

**ЗМІСТ**

Вступ	3
Застосовувані терміни та їх визначення	3
1.Огляд сучасного зарубіжного досвіду	4
2.Одиниця відбору	6
3.Зміст критеріїв	6
Джерела та література	9
Додаток 1	11

## ВСТУП

Аналітичне опрацювання і зберігання веб-ресурсів нині практикується у різних галузях діяльності людини. Так, у багатьох країнах світу нагромаджено досвід здійснення моніторингу і відбору інформації з мережі Інтернет в журналістиці<sup>1</sup>, науково-інформаційній<sup>2</sup> або політичній діяльності, органами слідства і контррозвідки. Накопичення та зберігання веб-ресурсів є актуальним завданням інформаційної спільноти<sup>3</sup>, оскільки, згідно з сучасними даними, щодня з'являються десятки мільйонів нових веб-сторінок і зникає близько семи мільйонів. Отже, можна зробити висновок щодо постійної втрати цінної для суспільства історичної фактографічної інформації.

Завдання відбору веб-інформації в архівній справі полягає у нагромадженні веб-ресурсів для можливого їх використання у подальшому як джерела ретроспективної інформації. Така діяльність є одним з основних напрямів роботи створеного в Україні Центрального державного електронного архіву України (далі – ЦДЕА України) і вона має здійснюватись на підставі певних критеріїв відбору.

**Мета розробки** – визначення критеріїв відбору веб-ресурсів як складової Національного архівного фонду і джерела комплектування ЦДЕА України.

### **Завдання розробки:**

- вивчення зарубіжного досвіду;
- визначення переліку та змісту критеріїв, на підставі яких може здійснюватися відбір веб-ресурсів в ЦДЕА України;
- представлення фрагментів рубрикатора, на підставі яких можуть бути сформульовані теми, за якими здійснюватиметься пошук веб-ресурсів та їх відбір для подальшого зберігання в ЦДЕА України.

## ЗАСТОСОВУВАНІ ТЕРМІНИ ТА ЇХ ВИЗНАЧЕННЯ

Поняття, що використовуються у цих Методичних рекомендаціях, вживаються у значеннях, визначених законами України («Про інформацію»<sup>4</sup>, «Про авторське право і суміжні права»<sup>5</sup>, «Про обов'язковий примірник документів»<sup>6</sup> «Про Національний архівний фонд та архівні установи»<sup>7</sup>) та чинним в Україні термінологічним стандартом з діловодства та архівної справи<sup>8</sup>. Інші терміни, що використані в Методичних рекомендаціях, уживаються в таких значеннях:

**веб-ресурс** – веб-сайт або сукупність веб-сайтів;

**веб-сайт** – сукупність програмних та апаратних засобів з унікальною адресою у мережі Інтернет разом з інформаційними ресурсами, що перебувають у розпорядженні певного суб'єкта і забезпечують доступ юридичних та фізичних осіб до цих інформаційних ресурсів та надають інші інформаційні послуги через мережу Інтернет;

**веб-документ** – документ в електронній формі, отриманий з мережі Інтернет у вигляді файлу будь-якого формату;

**архівна копія веб-сайта** – веб-сайт, що відібраний та скопійований для державного зберігання ЦДДА України.

### 1. Огляд сучасного зарубіжного досвіду

#### 1.1. США

**Інтернет-архів**<sup>9</sup> (англ. Internet Archive) – некомерційна організація, заснована у 1996 р. Архів зберігає копії веб-сайтів. Мета архіву – збереження культурно-історичних цінностей цивілізації в епоху використання Інтернет-технологій. Обсяг зберігаємої інформації – близько 5 петабайтів (близько 85 мільярдів веб-документів). Архів має юридичний статус бібліотеки. Інтернет-архів, як проект, зорієнтований на довгострокове зберігання інформації мережі всесвітнього павутиння для користування вченими, соціологами, журналістами, істориками, іншими користувачами, котрі могли б бути

зацікавленими у використанні інформації архіву в майбутньому. Відбір здійснюється періодичним скануванням глобальної мережі з каталогізацією, в основу якої покладено доменну ієрархію імен.

Проект «MINERVA»<sup>10</sup> – Інтернет-архів Бібліотеки Конгресу США Бібліотека здійснює відбір веб-документів, логічно поєднаних гіпертекстом та відповідають рубрикам визначеної тематики. Наприклад, тема – «Події», рубрики – «Вибори», «Олімпійські ігри», «11 вересня» тощо). Відбір здійснюється бібліотечною експертизою. Тематика формується за ознакою історичної цінності інформації.

## **1.2. Великобританія**

Національний архів Великобританії<sup>11</sup> зберігає інформацію основних урядових сайтів за таким рубриками:

- бізнес, промисловість, економіка, фінанси;
- культура;
- навколишнє середовище;
- уряд, політики та урядовці;
- охорона здоров'я;
- міжнародні справи та оборона;
- люди та суспільство;
- правопорядок, судочинство, право;
- транспорт, комунікації та технології;
- робота, навчання.

Національна бібліотека Великобританії<sup>12</sup> (British Library) також заявила про створення власного Інтернет-архіву, збір інформації буде здійснюватись з 10 тисяч сайтів Великобританії. В архіві будуть зберігатися веб-документи, що мають історичну та соціальну значущість.

Шість організацій Великобританії: the National Archives (Національний архів), the British Library (Британська бібліотека), the Joint Information Systems Committee – JISC (Комітет об'єднаних інформаційних систем), the National libraries of Scotland and Wales (Національна бібліотека Шотландії та Уельса),

the Wellcome Library (Публічна бібліотека), розробляють проект Інтернет-архіву і з цією метою заснували консорціум «the UK Web Archiving Consortium – UKWAC (Консорціум архівування Інтернету Великобританії)»<sup>13</sup>. Консорціум зорієнтовано на те, що кожен із членів для архівування веб-документів буде використовувати модифіковану версію програмного забезпечення Національної бібліотеки Австралії «PANDAS (Pandora Digital Archiving System)» та створювати архівний веб-ресурс, який відповідає їхнім інтересам та можливостям.

### **1.3. Австралія**

Проект архівування веб-ресурсів «PANDORA – Preserving and Accessing Networked Documentary Resources in Australia (Зберігання документальних ресурсів в Австралії з мережевим доступом)»<sup>14</sup> розпочато у 1996 р. Національною бібліотекою Австралії – National Library of Australia (NLA). На початку проекту відбір веб-документів здійснювався на основі особистих рішень бібліотекарів щодо цінності інформації. Згодом почали застосовувати автоматичний тематичний відбір інформації, наприклад, охорона здоров'я, вибори тощо. На підставі набутого досвіду розроблено інформаційну модель архіву «Logical Data Model (логічна модель даних)»<sup>15</sup>, яка представляє перелік собою близько декількох десятків тематичних рубрик з підрубриками. Зберігання веб-ресурсів здійснюється з орієнтацією на рекомендації проекту «OAIS – Open Archival Information System (Відкрита архівна інформаційна система)»<sup>16</sup>.

### **1.4. Ініціативи інших країн**

Проект восьми Європейських бібліотек «NEDLIB – Networked European Deposit Library (Депозитна бібліотека з мережевим доступом)»<sup>17</sup> спрямовано на депозитне архівування електронних публікацій.

Проект «Nordic Web Archive (Північний Веб-архів)»<sup>18</sup> національних бібліотек північних країн (Королівська бібліотека в Копенгагені, Бібліотеки Гельсінського університету, Національної та Університетської бібліотек Ісландії, Національної бібліотеки Норвегії та Королівської бібліотеки в

Стокгольмі) зорієнтовано на архівування веб-мережових електронних публікацій зі здійсненням відбору відповідно до бібліотечної класифікації.

### **1.5. Висновок**

За результатами опрацювання джерел огляду необхідно зазначити, що для усіх проектів типовим є згадування про невизначеність у вирішенні таких проблем, як врегулювання правових відносин у сфері інтелектуальної власності, доступ в процесі довгострокового зберігання, дублювання інформації іншими мас-медіа. Чітко формалізованих критеріїв відбору веб-ресурсів також не запропоновано, крім таких узагальнених, як тематика відбору та передбачувана цінність інформації для майбутнього.

## **2.ОДИНИЦЯ ВІДБОРУ**

Основною одиницею відбору є веб-сайт, який відповідає головним критеріям відбору і слугує для створення в ЦДЕА його архівної копії<sup>19</sup>. Водночас, за наявності розділів сайту, що містять значну кількість нерелевантної інформації і не відповідні критеріям, вони можуть вилучатися з нього.

Зважаючи на значущість для вітчизняного суспільства інформації, пріоритетному копіюванню підлягають веб-сайти державних органів та центральних державних установ України. На початковому етапі відбору, що має проводити ЦДЕА, доцільно до вирішення правових питань цієї діяльності архіву, формувати архівні копії веб-ресурсів зазначених та інших юридичних осіб, веб-сайти яких створені за державний кошт.

## **3. ЗМІСТ КРИТЕРІЇВ**

### **3.1. Диференціація критеріїв**

Критерії відбору веб-ресурсів можна поділити на дві групи.

Перша група – головні, визначальні критерії, до складу яких входить «тематика» та «значущість інформації для майбутніх споживачів».

Виокремлення зазначених критеріїв як пріоритетних зумовлено досвідом відбору веб-ресурсів за кордоном.

Друга – охоплює допоміжні критерії, які можуть застосовуватись, але не є визначальними і слугують лише додатковою мотивацією для відбору.

## **3.2. Головні критерії відбору**

**3.2.1. Критерій «тематика».** Відбір здійснюється відповідно до заздалегідь сформульованих тем і впродовж часу актуальності теми. Оскільки на сьогодні у дослідженнях вітчизняних істориків переважає проблематика, пов'язана з політичною та економічною історією України, то тематика відбору має зосереджуватись у цьому напрямі.

Водночас, така тематична обмеженість не повинна бути перепорою для відбору інформації з таких питань як, наприклад, «українці за кордоном», «відгуки за кордоном на події в Україні» тощо. Тематика може бути умовно поділеною на блоки: політика, економіка, наука, культура тощо. Фрагменти рубрикатора у вигляді тематичних блоків «Політика» та «Економіка» представлено у Додатку 1.

За наявності можливостей відбір може здійснюватись за всіма (або переважною більшістю) наведеними рубриками. Значне обмеження обсягів веб-ресурсів, що відбираються, може зумовлювати формулювання найактуальніших тем, наприклад, «вибори до Верховної Ради України», «вибори Президента України», «політична (економічна, фінансова, енергетична тощо) криза в Україні» та ін.

У разі розширення тематичного поля пошукових інформаційних запитів їх формулювання можна здійснювати за допомогою «Генеральної схеми класифікації документної інформації в систематичних каталогах державних архівів України»<sup>20</sup>.

## **3.2.2. Критерій «значущість інформації для майбутніх споживачів»**



Зазначений критерій фактично пов'язаний із попереднім, оскільки в інформаційній потребі головним все ж є формулювання теми (питання, проблеми).

Визначення можливих інформаційних потреб споживачів інформації у майбутньому ґрунтується не тільки на усталеному, традиційному інтересі всіх категорій споживачів (включаючи і дослідників) до так званих «сенсаційних новин», але й на аналізі нагромаджених фактів з історії життя країни та різних верств суспільства, можливості ретроспективного моніторингу подій, співмірної їм аналітики фактів і прогнозів. Рівень прогнозованої інформативності сайта формально залежить від рівня кумуляції інформації в ньому, його спрямованості як аналітичного джерела інформації, поза залежності від його профілізації як представника офіційної чи, навпаки, неофіційної (опозиційної або незалежної) точки зору. Чим вищі ці рівні у сайтів, тим вищі можливості їх відбору для формування їх архівних копій. З іншого боку, безперечно, значущість інформації сайта для майбутніх поколінь визначається на підставі експертних рішень персоналу, що відбирає веб-сайт, і у подальшому – експертно-перевірною комісією ЦДЕА України<sup>21</sup>.

### **3.3. Додаткові критерії**

#### **3.3.1. Критерій «джерело, з якого здійснено відбір»**

Сайт державного органу або центральної установи, який містить інформацію підтверджує її офіційність, має більшу міру вірогідності, що оцінюється позитивно. Повідомлення, вміщене на сайті пересічної юридичної або фізичної особи, може мати значно меншу тривалість існування й більшу міру унікальності порівняно з попередніми видами джерел, а значить бути більш цінним.

Сайти як джерела інформації, що вже відзначалося вище, можуть мати різні рівні кумуляції інформації та вмісту аналітичної (фактологічної) інформації. З другого боку, міра кумуляції інформації з теми має узгоджуватись з такою змістовною характеристикою, як повнота інформації.

Тобто, необхідно здійснювати порівняльний аналіз веб-ресурсів щодо міри презентативності інформації з теми на різних сайтах, що впливатиме на їх вибір як джерела для відбору.

Слід також врахувати, що на різних сайтах можуть вміщуватися різні версії фактів, їхні коментарі, аналітичні матеріали, що зумовлює необхідність копіювання ресурсів, які містять різні (іноді полярно протилежні) оцінки фактів. З цієї точки зору найбільшу цінність для відбору представляють сайти аналітики новин, на яких часто спостерігається дублювання інформації, що вже зафіксована (ефект поглинання інформації). Разом з тим, аналітико-синтетичні операції щодо вилучення дублюючих частин сайтів можуть займати багато часу. Тобто копіювання сайтів без їх аналітичного опрацювання буде охоплювати значно менше часу, але у цьому разі є більша ймовірність зберігання значних обсягів нерелевантної інформації.

### **3.3.2. Критерій «міра достовірності інформації»**

Міра достовірності інформації на рівні відбору та у процесі колегіальної експертизи цінності визначається за результатами контент-аналізу сайту як джерела<sup>22</sup>. Зазначений критерій може мати різні ступені: від сприйняття інформації як недостовірної до наявності її офіційного підтвердження. Оскільки нагромадження інформації з теми відбувається постійно, то веб-ресурси, які містять окремі повідомлення, спростовані офіційно, не вилучають, а залишають із наступним спростуванням. Особливу увагу слід приділити так званим «чуткам» або «пліткам». Ця інформація зникає швидше, але саме вона є атрибутом соціально-комунікаційного тла епохи і її наявність не дає можливості представити історію за офіційними виданнями як суцільну пряму лінію. Важливим є збереження вихідного повідомлення, яке має т. зв. “сенсаційний характер”. Водночас, сенсаційні новини, окрім підтверджених, у подальшому, як правило, або зникають без сліду і про них забувають, або, найбільш гучні, – спростовуються і розглядаються у судах (за наявністю позовів).

### **3.3.3. Критерій «визначеність авторства»**

Безперечно, визначеність автора створює більш надійне фактичне підґрунтя ресурсу. Водночас, на сайтах новин часто відсутнє зазначення авторства інформації, а має місце посилення на сайт або інформаційну агенцію, звідки її взято. Відтак, відбір здійснюється незалежно від авторства повідомлення, вміщеного на якомусь сайті.

## Джерела і література

- <sup>1</sup> Снурнікова Ю. М. Ефективне використання електронних документів мережі Інтернет в тележурналістиці: Автореф. дис. ... канд. філол. наук. / Київ. нац. ун-т ім. Тараса Шевченка; Ін-т журналістики. – К., 2005. – 20 с.
- <sup>2</sup> Копанєва В.О. Формування фонду мережевих ресурсів у науковій бібліотеці (90-ті роки ХХ – поч. ХХІ ст.) Автореф. дис. ... канд. іст. наук/НБУВ. – К., 2008. – 20 с.
- <sup>3</sup> [http://www.rocit.ru/public/index.php3?path=ex\\_uno](http://www.rocit.ru/public/index.php3?path=ex_uno)
- <sup>4</sup> Про інформацію: Закон України від 2 жовтня 1992 року N 2657-XII // Відомості Верховної Ради України. — 1992. — N 48. — С. 650.
- <sup>5</sup> Про авторське право і суміжні права: Закон України від 23 грудня 1993 року N 3792-XII // Відомості Верховної Ради України. — 1994. — N 13. — С.64.
- <sup>6</sup> Про обов'язковий примірник документів: Закон України// ВВР. – 1999. - № 22-23. – Ст.199 (із змінами).
- <sup>7</sup> Про Національний архівний фонд та архівні установи: Закон України// Офіційний вісник України. – 2002. – № 2. – Ст. 43.
- <sup>8</sup> ДСТУ 2732: 2004. Діловодство і архівна справа. Терміни та визначення понять / Держспоживстандарт України. — К., 2005. — 33с.
- <sup>9</sup> The Internet Archive./<http://web.archive.org/collections/web/>
- <sup>10</sup> <http://www.rlg.org/preserv/diginews/diginews5-2.html#feature1>
- <sup>11</sup> <http://www.nationalarchives.gov.uk/>
- <sup>12</sup> British Library Collection Development Policy  
<http://www.bl.uk/about/policies/collections.html>
- <sup>13</sup> UK Web Archiving Consortium (UKWAC) <http://www.webarchive.org.uk/>
- <sup>14</sup> <http://pandora.nla.gov.au/pandora/>
- <sup>15</sup> <http://www.nla.gov.au/pandora/ldmv2.html>
- <sup>16</sup> <http://ssdoo.gsfc.nasa.gov/nost/isoas/>
- <sup>17</sup> <http://www.kb.nl/coop/nedlib/>
- <sup>18</sup> <http://nwa.nb.no>
- <sup>19</sup> Про створення та зберігання архівних копій веб-сайтів: Методичні рекомендації/ Центральный державний електронний архів України (схвалено Нормативно-методичною комісією Держкомархіву України 26.05.2009 р., протокол № 2).
- <sup>20</sup> Генеральна схема класифікації документної інформації в систематичних каталогах державних архівів України / Держкомархів України, УНДІАСД. Розробники: Н.М. Христова (кер. розробки), Л.В. Андрієвська (відп. виконавець) та ін. – К., 2006. – 137 с.
- <sup>21</sup> Про створення та зберігання архівних копій веб-сайтів: Методичні рекомендації/ Центральный державний електронний архів України, пункт 2.3.
- <sup>22</sup> Там само, пункт 4.4.2.

## Додаток 1

## Критерії відбору веб-ресурсів

№№ пп	Рівень відповід- ності темі	Значущість інформації	Визначеність авторства	Джерело, з якого відбирається інформація	Міра вірогід- ності інфор- мації	Міра поширення інформації	Міра кумуляції і концентрації інформації
1	2	3	4	5	6	7	8
1	Повна відповід- ність	Загально- державна	Визначено	Сайт пересічної юридичної або фізичної особи	Підтверджена офіційно	Тільки на одному сайті юридичної або фізичної особи	Аналітичний огляд фактів з ретроспек- тивним аналізом
2	Часткова відповід- ність	Регіональна	Псевдонімічне	Сайт електронного видання	Невизначена, але цілком вірогідна	Поширена на ба- гатьох сайтах юридичних і фізичних осіб	Описання факту або дайджест фактів з аналітичним комента-рем
3	Дотична до теми	Місцева	Не визначено (анонімне)	Сайт держа- вного органу, центральної установи	Невизначена і сумнівна	Продубльована в електронному виданні	Дайджест фактів без коментарів
4						Є в друкованому виданні	Описання факту