

УДК 681.3+519.83

ФУРАШЕВ О.В., керівник проектів ТОВ “Арїон – Консалтинг”**ПРО ОСОБЛИВОСТІ РОБОТИ З ЕЛЕКТРОННИМИ
ІНФОРМАЦІЙНИМИ РЕСУРСАМИ**

Анотація. Щодо визначення напрямів зменшення трафіку передачі даних при збереженні обсягу переданої інформації.

Аннотация. Об определении направлений уменьшения трафика передачи данных при сохранении объема передаваемой информации.

Summary. As to determination of main directions in decreasing of data transmission traffic when the volume of transmitted information is saved.

Ключові слова: дані, формат даних, електронний інформаційний простір.

Інформаційний простір, з якого нині людина дістає інформацію значимого змісту, необхідного для ухвалення рішень і здійснення практичної діяльності, трансформується. Змінилися як засоби доставки інформації, так і сам формат. Все більшу частину інформації, до якої можна отримати доступ самостійно, людина обробляє в електронному вигляді. Такий інформаційний простір, в якому людина самостійно задовольняє свої інформаційні потреби за допомогою отримання і обробки інформації в електронному вигляді, будемо називати сучасним електронним простором. На сьогодні в ньому, на нашу думку, можна виділити наступні властивості:

- інформація представлена в електронному вигляді;
- доступність каналів зв'язку і засобів доступу до них;
- оперативність пошуку та отримання інформації;
- доступність великої кількості інформаційних джерел;
- отримання авторської інформації без редакторських правок та скорочень;
- кожен учасник виступає як постачальник, так і споживач інформації.

До недоліків сучасного електронного інформаційного простору, на думку автора, слід віднести:

- значні часові витрати на отримання саме потрібної інформації, незважаючи на оперативність пошуку та доступність інформаційних ресурсів;
- постійно зростаючі вимоги до засобів доступу до інформаційних ресурсів та каналів зв'язку.

Склалася суперечлива ситуація: з одного боку, збільшилось число інформаційних джерел, час доступу до них скорочується, збільшується продуктивність засобів зв'язку та обробки даних, а з іншого боку, ресурсоемність отримання потрібної інформації не зменшилась.

Метою роботи є визначення напрямів роботи з електронними інформаційними ресурсами, які зменшать обсяг та ресурсоемність обробки даних. Для досягнення цієї мети розглянемо структуру сучасного інформаційного простору і основні процеси щодо взаємодії його учасників.

Електронний інформаційний простір складається з наступних основних компонентів:

- дані – інформація, що перетворена в електронний вигляд і закодована способом, який дозволяє передавати її електронними каналами зв'язку та відтворювати в зрозумілій для людини формі;

• людина – суб’єкт сучасного електронного інформаційного простору, об’єктом його діяльності є дані. Залежно від її відношення до інформації вона може виконувати наступні ролі:

- автор – породжує інформацію та розміщує її в інформаційному просторі;
- адресат – споживає інформацію. Він отримує, обробляє та приймає на її основі рішення. По відношенню до автора адресат може бути:
 - відомим – авторові відомі одержувачі його інформації;
 - анонімним – автор адресує свою інформацію певній групі людей;

• обладнання візуалізації і доступу (далі – ОВД). Його основні функції: 1) перетворювати дані у вид, зрозумілий людині, тобто візуалізувати; 2) забезпечити підключення до комунікаційних пристроїв мережі, прийом/передачу даних через мережу. На сучасному етапі розвитку ОВД обслуговує одну людину, є його власністю. До ОВД відносяться ПК, планшети, смартфони, телефони і т.п.;

• мережа – система, що складається з комунікаційних пристроїв і ліній зв’язку. Основною її функцією є передача даних між комунікаційними пристроями.

• майданчик – сервер з високою мірою доступності, постійно підключений до мережі. Його функція – збереження та передача розташованих на ньому даних без істотних обмежень за часом та кількістю запитів на отримання даних.

Схема взаємодії між основними компонентами в сучасному електронному інформаційному просторі наведена на Рис 1.

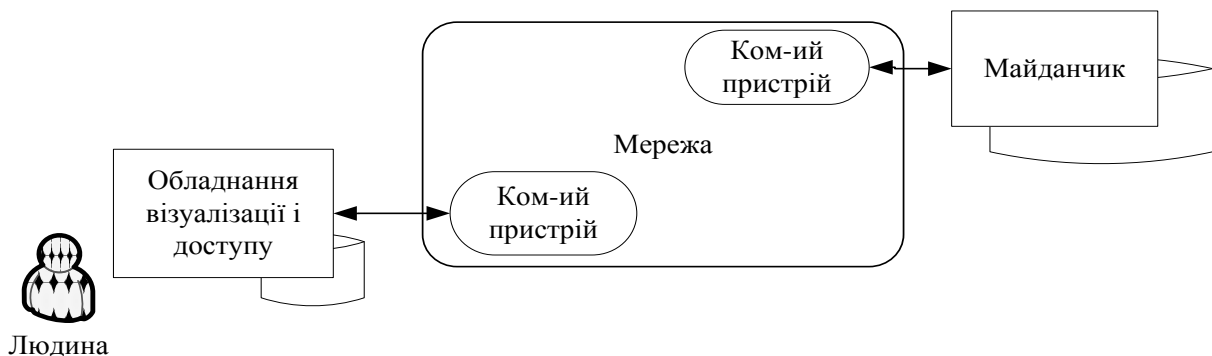


Рис. 1. Схема взаємодії між основними компонентами в сучасному електронному інформаційному просторі

Із схеми видно, що в електронному просторі людина присутня через посередника (далі – ОВД), їх може бути декілька. Для забезпечення доступності даних вводиться додаткова ланка – Майданчик.

Робота людини в сучасному електронному інформаційному просторі зводиться до трьох основних процесів:

- 1) підготовка та публікація даних;
- 2) отримання даних та їх обробка;
- 3) пошук даних.

Розглянемо перші два процеси. На рис. 2 наведена схема узагальненого процесу підготовки та публікації інформації.

Початок процесу публікації визначається необхідністю передавати адресатам будь-яку інформацію. Це може бути як звичайна відповідь, питання у форумі, так і детальний звіт, документ. Інформація готується автором в одному з поширених форматів в інформаційному просторі: в текстовому, графічному і т.п.



Рис 2. Схема узагальненого процесу підготовки та публікації інформації

Перетворена інформація набуває вигляду, що читається машиною, яку надалі будемо назвати даними. На кроці “Публікація даних” дані стають доступними адресатові шляхом їх розміщення на майданчику, через який він зможе їх отримати. Після того як дані розміщені на майданчику, автор повідомляє адресатові адресу місця розташування даних, зазвичай, у вигляді URL.

На Рис. 3 наведена схема узагальненого процесу отримання даних, який розпочинається з моменту, коли адресатові стає відома адреса місця розташування даних. За супутньою інформацією адресат приймає рішення про отримання даних і виконує наступний крок процесу. ОВД намагається знайти дані за вказаним місцем розташування. Знаходить майданчик, який містить дані, встановлює з ним зв’язок і дістає дані. Після їх отримання дані візуалізуються, людина вивчає їх і приймає рішення щодо того, наскільки отримані дані є значимими.

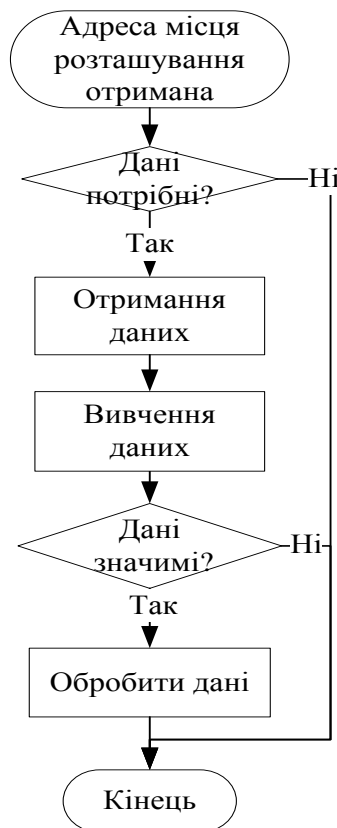


Рис 3. Схема узагальненого процесу отримання даних

Обидва процеси мають лише причинно-наслідковий зв’язок. Так, процес отримання даних не може бути розпочатий, поки не буде виконаний процес їх підготовки та публікації, але зворотне не вірно – завершення процесу публікації не завжди призводить до початку процесу фактичного отримання даних. Для забезпечення передачі

результатів процесу публікації до процесу витягання даних був введений проміжний елемент – майданчик. Він не лише виконує функції буфера, а й доповнює іншими, спрямованими на інтеграцію розміщуваних даних у загальний масив, що міститься на майданчику. Як наслідок, дані втрачають самостійну форму і, як результат, адресат отримує з майданчика:

- окрім авторської інформації, додаткову інформацію, як, наприклад: елементи оформлення, навігаційні або рекламні блоки і т.п.;
- дані можуть бути перетворені та передані в іншому форматі.

Це, у свою чергу, веде до збільшення обсягу даних, що передаються, без зміни авторської інформації. Додатковий об’єм даних призводить до навантажень на мережу і ОВД, що збільшує ресурси, потрібні для передачі і обробки даних. Додаткова інформація збільшує також і навантаження на людину, їй необхідно самостійно виділити і знайти авторську інформацію.

Оцінимо міру впливу цих чинників на ресурсоемність процесу отримання даних.

Оцінка міри зроблена на основі тексту Податкового Кодексу України [1]. Співвідношення обсягів документа наведено в Таблиці 1. Для оцінки було взято найпопулярніші формати представлення інформації, що використовуються при передачі даних у мережі Інтернет. Параметр оцінки часу на передачу даних визначався для швидкості 6 КБ/з найбільш характерною для мобільних мереж покоління 2,5G [2]. Документ був представлений у форматах:

- оригінальний html у вигляді, в якому він розміщений на сайті Верховної Ради України, включаючи елементи оформлення і навігації;
- оригінальний html без графічних елементів, з обсягу документа були виключені усі графічні елементи, представлений лише текст;
- оптимізований html. У документі були залишені тільки структуротворні теги;
- текстовий у простому кодуванні. Документ був перетворений в текст з розбиттям на абзаци без будь-якого шрифтового виділення;
- текстовий, в кодуванні UTF8. Документ у текстовому форматі був перетворений в кодування міжнародного формату – UTF8;
- у форматі текстового процесора MS Word, версії 97-2003;
- PDF.

Таблиця 1.

Формат	Обсяг даних, КБ	% до оригінального	Співвідношення з текстовим представленням	Оцінка часу на передачу даних, с.
Оригінальний html	120	100 %	2,79	20
Оригінальний html без графічних елементів	70	58 %	1,63	12
Оптимізований html	44	63 %	1,02	7
Текстовий у простому кодуванні	43	61 %	1,00	7
Текстовий у кодуванні UTF8	78	111 %	1,81	13
MS Word, версії 97-2003	111	159 %	2,58	19
PDF	330	471 %	7,67	55

З наведеного видно, що, не змінюючи авторського тексту, а тільки за рахунок формату представлення, обсяг даних змінюється. Якщо взяти за основу представлення документа в текстовому форматі, то із зміною формату відбувається збільшення їх

обсягу, а це, у свою чергу, призводить до збільшення ресурсів, необхідних для отримання інформації.

Оцінка міри впливу додаткових даних наведена на прикладі статті про Україну з Wikipedia [3] та 10 опублікованих новин на сайті rodnosti.ua. Дані перетворювались в такі види без спотворення інформаційного наповнення:

- оригінальний. Дані оброблювались в тому вигляді, в якому були передані з сайту;
- оптимізований. З даних була видалена уся службова інформація, рекламні і навігаційні блоки, ілюстрації збережені;
- дані в wiki-розмітці. Представлення даних у цьому форматі відображає авторський текст, яким його підготував і опублікував би автор, включаючи ілюстрації.

Таблиця 2.

Вид даних	Обсяг даних, КБ	% до оригінального	Співвідношення з текстовим представленням	Оцінка часу на передачу даних, хв.
Оригінальний	6198	100%	5,85	17
Оптимізований	1337	22%	1,26	4
Дані в wiki-розмітці	1060	17%	1,00	3

З Таблиці 2 видно, що обсяг даних, який надходить адресатові, істотно перевищує об'єм даних, що передавав автор. Тобто нині збільшення обсягу даних в сучасному інформаційному просторі відбувається завдяки збільшенню обсягу додаткової інформації та використанню не найоптимальніших форматів її представлення. Причому обидва чинники значною мірою впливають на ресурсоемність, що є причиною, через яку компанія Apple відмовилася включати у свої платформи iPhone і iPad підтримку даних у форматі flash [4]. У цей формат перетворюють в основному додаткові дані (рекламні блоки, навігаційні елементи тощо).

Статистика на сайті [5] підтверджує, що в структурі даних, що передаються в мережі Інтернет істотну частку займають додаткові дані. Це веде до підвищення вимог до продуктивності УВЗ та каналів зв'язку та врешті – до збільшення його витрат.

Або користувачам необхідно змінити спосіб роботи в інформаційному просторі, таким чином, щоб знизити обсяг додаткових даних та отримувати їх у менш ресурсоемному форматі. Виходячи з розглянутих процесів роботи і взаємодії учасників в сучасному електронному інформаційному просторі, визначимо напрями, в яких слід змінювати принцип роботи:

- скористатися одним із спеціалізованих сервісів, який з даних, що передаються, видаляє додаткові блоки і перекодує їх у більш компактний формат. Прикладом такого сервісу може бути Turbo Opera [6];

- змінити поведінку ОВД таким чином, щоб він не направляв запит на додаткові блоки даних. Суть у тому, що у найпоширенішому форматі представлення даних html різні дані зберігаються в окремих файлах і потребують від ОВД окремих запитів на їх отримання. Це досягається як допоміжними програмами, названими “блокувальниками реклами”, наприклад Adblock, так і зміною налаштувань самого браузера;

- по можливості запрошувати інформацію в компактнішій формі. Наприклад, багато сайтів мають так звані “мобільні” версії, оптимізовані під УВС з обмеженими ресурсами. Такі версії сайтів надають в основному статичну інформацію і тому містять меншу кількість додаткової інформації, дані представляються у компактнішому форматі.

Запропоновані принципи можуть застосовуватись нині і вносять зміни тільки на рівні УВС, не зачіпаючи інших учасників інформаційного простору, але не усувають основну причину проблеми. Для більш комплексного вирішення проблеми необхідно зменшити міру участі майданчика у взаємодії між автором і адресатами. Цього можна досягти таким чином:

- розташувати майданчик на стороні автора або адресатів. Необхідність майданчика обумовлена тим, що адресати не постійно доступні. Із здешевленням тарифів використання мережі час присутності в мережі учасників зростає. Тому від проміжної ланки в організації процесу можна відмовитися і організувати обмін даними безпосередньо між її учасниками. У цьому напрямі розвиваються так звані P2P мережі, на цьому принципі побудовані деякі комунікаційні програми, наприклад, Skype;

- публікувати інформацію в кінцевому форматі для адресата. Велика частина інформації нині публікується у проміжному виді, обумовленому майданчиком;

- публікувати інформацію у найкомпактнішому форматі – текст з елементами структуризації. Наприклад, у форматі wiki-розмітка. Цієї ж мети досягає формат представлення XML. Тут відзначимо, що wiki-розмітка текстів менш трудомістка і виконується вручну. Представлення текстів у вигляді XML-формату, навпаки, досить трудомістке, візуалізація вимагає додаткових даних і є ресурсоемним завданням.

Як видно, вимоги до продуктивності ОВД, ширини каналу доступу нині є у більшій мірі штучними і створені за рахунок збільшення обсягу додаткових даних. За рахунок зміни налаштувань ОВД, спрямованих на зменшення запитів отримання допоміжних даних, можливе зменшення обсягу отримуваних даних і тим самим зменшення вимог до пропускнує спроможності каналу і продуктивності пристроїв, що використовуються. Основні способи зменшення частини непотрібного обсягу даних вимагають зменшення міри участі майданчиків у процесі публікації і отримання даних. Для цього необхідно змінювати вже не налаштування ОВД, а способи публікації даних, їх обробки і отримання.

Надання публікованої інформації самостійної форми, взаємодія безпосередньо між автором і адресатами, представлення інформації в текстовому вигляді з елементами її структуризації – це ті напрями, які зменшать обсяг допоміжних даних та ресурсоемність їх обробки. На думку автора, ці напрями потрібно розвивати надалі, оскільки основною їх рушійною силою є вимоги самих учасників. Зазначене вимагатиме зміни організації окремих кроків процесів і адаптації до їх підтримки ОВД.

Використана література

1. Податковий кодекс України : Закон України. – Режим доступу : [//www.zakon1.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=2755-17](http://www.zakon1.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=2755-17)
2. Википедия. Поколения мобильной телефонии. – Режим доступу : [// www.ru.wikipedia.org/wiki](http://www.ru.wikipedia.org/wiki)
3. Википедия. Украина. – Режим доступу : [// www.ru.wikipedia.org/wiki](http://www.ru.wikipedia.org/wiki)
4. Джобс не пустит Flash в продукты Apple. – Режим доступу : [// www.lenta.ru/news/2010/04/30/jobs](http://www.lenta.ru/news/2010/04/30/jobs)
5. Interesting stats. “Average bytes per pages by content type”, “Avg individual resource response size”, “Pages with the most javascript”, “Pages using flash”. – Режим доступу : [// www.httparchive.org/interesting.php](http://www.httparchive.org/interesting.php)
6. Википедия. Opera Turbo. – Режим доступу : [//www.ru.wikipedia.org/wiki/Opera_Turbo](http://www.ru.wikipedia.org/wiki/Opera_Turbo)

~~~~~ \* \* \* ~~~~~