

Бондарчук С. В.

(0000-0003-0624-9782)

Центр воєнно-стратегічних досліджень Національного університету оборони України імені Івана Черняхівського, Київ

## Обґрунтування підходу щодо підготовки даних до первинного завантаження для проведення тестування інформаційних систем управління оборонними ресурсами

**Резюме.** У статті розглянуто підхід щодо підготовки тестових даних до первинного завантаження в базу даних для проведення тестування інформаційних систем управління оборонними ресурсами. На основі проведеного аналізу визначено проблемні питання під час перетворення даних і шляхи їх вирішення. Запропоновано інструментарій та визначено етапи обробки і перетворення неструктурованої та/або частково структурованої інформації в структуровану.

**Ключові слова:** управління оборонними ресурсами; підготовка даних для тестування; перетворення даних; структурування даних; структуровані дані.

**Постановка проблеми.** Україна, зокрема і збройні сили проходять етап становлення в нових ринкових відносинах. Ефективне використання оборонних ресурсів залежить від швидкого збору, обробки інформації, і як результат – отримання повної і достовірної інформації про стан і наявність ресурсів для правильного прийняття ефективних рішень [1]. Централізація управління і потреба в розумінні кількісного та якісного стану наявних ресурсів спонукає до автоматизації і впровадження інформаційних систем обліку та управління оборонними ресурсами. Підґрунтям для ефективної роботи таких систем є бази даних та інструментарій щодо управління ними.

На етапі створення кожної бази даних виникає потреба в разовому (первинному), але досить масштабному (об'ємному) процесі підготовки даних шляхом перетворення і завантаження великих обсягів неструктурованої (частково структурованої) інформації із зовнішніх джерел для проведення тестування інформаційної системи (ІС). Одним із завдань Оборонної реформи ЗС України є створення *єдиної інформаційної системи управління оборонними ресурсами* (Defense Resources Management Information System – DRMIS), яка передбачає інтеграцію ІС управління особовим складом, управління організаційною структурою, управління оборонним плануванням, матеріально-технічним, медичним та іншими видами забезпечення, управління закупівлями, управління майном, фінансами та бюджетом, адміністративно-господарською діяльністю [2–5]. Отже, одноразова потреба завантаження великого об'єму даних для

конкретної ІС, в масштабі воєнного відомства перетворюється у багаторазову.

До розробника кожної ІС попередньо надходить велика кількість різномірної та непослідовної інформації, яка перебуває в різних стандартах і форматах. Якість підготовки даних і їх формати дуже сильно залежать від людського фактора, потреб і вподобань користувачів на етапі попереднього формування. Використання різномірних джерел даних потребує певного розуміння правил щодо їх поєднання у базі даних ІС, які не дадуть втратити структури зв'язків та суть інформації. Сьогодні також має враховувати існуючі матеріальні, фінансові обмеження, щоб вчасно відпрацювати відповідні етапи життєвого циклу ІС. Тому нагальним практичним завданням є скорочення витрат на підготовку даних на етапі тестування ІС із забезпеченням якості й повноти первинного наповнення бази даних.

**Аналіз останніх досліджень і публікацій.** На сьогодні існують спеціалізовані програмні продукти з аналізу текстів (Textual Analysis Computing Tools (ТАСТ), HTML Batch Editor, Словогрыз, ClearText Reader's Edition, xReplacer, xMarkup, TODOS), деякі з них можуть будувати зв'язки. Однак для роботи з ними потрібні чи спеціальні знання, чи замовлення вартісної послуги і залучення відповідного виконавця [6–8]. Саме тому, для підготовки даних до завантаження та проведення тестування новостворюваних ІС, здебільшого застосовують існуючий загальнодоступний інструментарій програм роботи з текстом, подібних до Notepad та Microsoft Office. Зазвичай цей інструментарій використовують

різні розробники ІС незалежно один від одного. Кожний розробник під час створення конкретної ІС одноразово знаходить свій унікальний спосіб перетворення даних, який не розголошується. Отже, обґрунтування методичного забезпечення щодо застосування загальноживаного інструментарію для первинної підготовки даних є актуальним завданням, відпрацювання якого дасть змогу задовольнити значну частину вимог щодо попередньої підготовки даних, максимально оптимізувати і, за необхідності, повторювати цей процес для проведення тестування різних ІС.

**Метою статті є формалізація процесу підготовки даних до первинного завантаження в базу даних із застосуванням загальноживаного інструментарію для проведення тестування інформаційних систем управління оборонними ресурсами.**

**Викладення основного матеріалу.** Під час первинного завантаження даних в базу даних, важливо усвідомлювати процеси обробки та перетворення даних, а також

приблизно розуміти структуру первинних (початкових) даних і чітко розуміти структуру кінцевих даних та їх призначення. За суттю, процес первинного завантаження можна віднести до кібернетичної трактовки роботи із сигналами та процесами [9]. Кібернетична трактовка цього поняття передбачає відмову від вивчення фізичних особливостей як первинного процесу, так і носія сигналу. Сигнал ототожнюється з інформацією про фізичні характеристики процесу (об'єкта, явища), який досліджується. До того ж ураховується, що з різних причин початкові дані не містять усієї інформації про певний об'єкт даних, а з іншого боку, початкова інформація може містити сторонню інформацію. На інформаційне наповнення первинних даних впливають способи їх обробки, зберігання, помилки та інші чинники.

Загалом кібернетичний блок (“чорний ящик”) – це блок, для якого встановлено пов'язані причинно-наслідковим відношенням вхідні та вихідні сигнали (рис. 1).

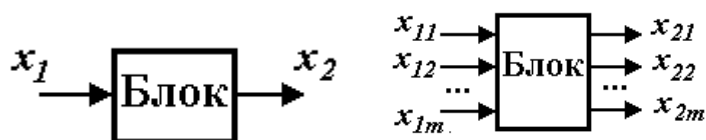


Рис. 1. Одноканальний та багатоканальний кібернетичні блоки

Для отримання необхідного вихідного сигналу блоку  $x_2(t)$ , внутрішні процеси змінюються та повторюються залежно від вхідного сигналу  $x_1(t)$ , який несе інформацію, проте важливо не втратити чи спотворити інформацію. У класичному визначенні кібернетичного блоку відсутня згадка про фізичну природу процесів усередині блоку, що і визначається терміном “чорний ящик”. Залежно від числа вхідних і вихідних сигналів розрізняють одноканальні блоки, тобто блоки з одним входом і одним виходом, і багатоканальні – з декількома вхідними та вихідними сигналами. Для спрощення, процес перетворення даних розглянемо як одноканальний кібернетичний блок.

Для опису кібернетичного блоку використовується одна з форм аналітичного опису зв'язку вхідних і вихідних сигналів. Для одноканального блоку такий опис може бути у вигляді алгебраїчного чи трансцендентного рівняння  $x_2 = f(x_1)$ , де  $f(\cdot)$  – функція перетворення даних. Для завантаження даних, їх необхідно упорядкувати, і цей процес виконується в одноканальному блоці, де вони

перетворюючись переходять до бази даних. Операції з обробки даних необхідно виконувати за певними правилами. Ці правила визначаються структурою (моделлю) даних [10, 11]. За допомогою моделі даних можуть бути представлені об'єкти предметної області та взаємозв'язки між ними [12].

Зазвичай використовується реляційна, або подібна модель даних, яка ґрунтується на добре опрацьованій теорії відносин (*Relation* – відношення). Під час проектування такої бази даних застосовуються чіткі методи, побудовані на нормалізації відносин. Дані в реляційній моделі представляються у вигляді таблиць, які називаються відносинами, де стовпці таблиці – елементи даних, а рядки – записи. Допустимими відносинами в реляційній моделі є ті, які задовольняють умові: на перетині будь-якого рядка і будь-якого стовпця в таблиці має бути одне, а не безліч значень. Сукупність значень зв'язаних елементів даних утворює запис даних.

Кожен об'єкт в базі даних характеризується деяким набором атрибутів (наприклад, об'єкт людина може мати такі атрибути як прізвище, адреса і, можливо, ідентифікаційний номер). Унікальне значення,

яке приймає атрибут об'єкта, дає змогу ідентифікувати значення, які приймають інші елементи даних цього ж об'єкта (наприклад, знаючи ідентифікаційний номер людини, можна дізнатися прізвище, адресу, дату народження). Елементи даних, за якими можна визначити інші елементи даних, називаються ключовими.

Не кожні відносини матимуть первинний ключ у вигляді єдиного атрибуту, але можна використовувати деяку комбінацію атрибутів, які взяті разом однозначно

ідентифікуватимуть значення у відносинах. Тоді, первинний ключ називатиметься складеним первинним ключем.

Первинна інформація, як правило, зберігається в частково чи повністю неструктурованому вигляді у текстових форматах, які можуть оброблятися за допомогою загальноживаних програмних продуктів Microsoft Office – Word чи Excel. Як приклад, інформація може мати такий вигляд (рис. 2):

№ з/п	П. І. Б.	Рік народження
1	Сидоренко І.І.	1960 р.
2	Петренко Іван Петрович	68 р.н.
.....	.....	.....
5125	Яценко Степан	1972

№ з/п	П. І. Б.
1	Сидоренко І.І. -1960 р.
2	Петренко Іван Петрович - 68 р.н.
.....	.....
5 25	Яценко Степан - 1972

Математично це можна відобразити так:

Атрибут	Заголовок 1	Заголовок 2
k <sub>1</sub>	a <sub>1,1</sub> a <sub>1,2</sub> .	d <sub>1,1</sub> d <sub>1,2</sub>
k <sub>2</sub>	a <sub>2,1</sub>	d <sub>2,1</sub> d <sub>2,2</sub>
.....	.....	.....
k <sub>n</sub>	a <sub>n,1</sub> a <sub>n,2</sub> a <sub>n,3</sub>	d <sub>n</sub>

Атрибут	Заголовок
k <sub>1</sub>	a <sub>1,1</sub> a <sub>1,2</sub> . d <sub>1</sub>
k <sub>2</sub>	a <sub>2,1</sub> a <sub>2,2</sub> . d <sub>2</sub>
.....	.....
k <sub>n</sub>	a <sub>n,1</sub> a <sub>n,2</sub> a <sub>n,3</sub> d <sub>n</sub>

де a, d – дані (будь-які дані)

k – атрибути (k<sub>i</sub>≠k<sub>j</sub>, тільки один для даної таблиці)

Рис. 2. Вигляд даних до перетворення

З прикладу видно, що дані можуть бути як у різних стовпцях, так і в одному; особливий випадок, коли вони являють собою однотипні дані (наприклад, прізвище, ім'я, по батькові). Загалом *опис та обробка текстів, таблиць* – це морфологічний, лексичний і синтаксичний аналіз фраз для виділення інформації щодо об'єктів (людей, географічних понять, наприклад, адреси

організацій, продукти, матеріали та ін.), їх взаємозв'язків. Розуміння цього надалі дасть змогу провести структурування виявлених фактів і зв'язків у вигляді реляційних таблиць, які потім можуть бути завантажені до бази даних [12].

Вигляд, який мають дані після структурування наведено на рис. 3.

Атрибут	Заголовок 1	Заголовок 2	Заголовок 3	.....	Заголовок n
1	Сидоренко	І.	І.	.....	1960
2	Петренко	Іван	Петрович	.....	68
.....	.....	.....	.....	.....	.....
5125	Яценко	Степан	.....	.....	1972

Математично це можна відобразити так:

Атрибут	Заголовок 1	Заголовок 2	Заголовок 3	.....	Заголовок n
k <sub>1</sub>	a <sub>1,1</sub>	a <sub>1,2</sub>	a <sub>1,n</sub>	.....	d <sub>1</sub>
k <sub>2</sub>	a <sub>2,1</sub>	a <sub>2,2</sub>	a <sub>2,n</sub>	.....	d <sub>2</sub>
.....	.....	.....	.....	.....	.....
k <sub>n</sub>	a <sub>n,1</sub>	a <sub>n,2</sub>	a <sub>n,n</sub>	.....	d <sub>n</sub>

Рис. 3. Вигляд даних після перетворення (кінцевих даних)

З рис. 3 видно, що кожне поле таблиці знаходиться окремо на перетині горизонталі та вертикалі, кожний запис має тільки один атрибут і за його допомогою може бути ідентифікований однозначно.

У процесі завантаження та перетворення дані проходять основні стадії, кожна з яких реалізується у вигляді окремої фази процесу перетворення даних (табл. 1).

Фази процесу перетворення даних

№	Назва	Опис
1	Витягування	Стадія витягування даних з джерела і завантаження їх в допоміжну область
2	Виявлення помилок	Дані проходять перевірку на відповідність специфікаціям і потенційну можливість завантаження в базу даних
3	Перетворення	Дані групуються і перетворюються згідно з моделлю даних
4	Розподіл	Дані розподіляються на декілька потоків (формується декілька таблиць), залежно від способу, яким вони мають бути завантажені до бази даних
5	Вставка (завантаження)	Підготовлені дані поступають до бази даних

Фази 1–3 є найбільш громіздкими в обробці та аналізі, потребують значної уваги та одноманітного повторення операцій до досягнення повністю структурованого вигляду даних. Фази 4 та 5 виконуються стандартними засобами системи керування базою даних.

Далі розглянемо перші три фази більш детально. На кожній стадії слід визначитися з початковими джерелами (Word чи Excel), допоміжною областю (Word, Excel та TXT) і кінцевим приймачем – базою даних (рис. 4).

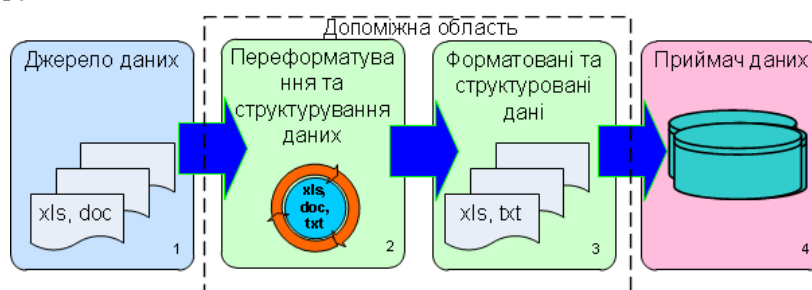


Рис. 4. Кроки процесу перетворення даних

Excel найбільш підходить для обробки інформації у форматі чисел, зазвичай, більша частина інформації в джерелах, знаходиться саме в цьому форматі. Також табличний процесор Excel можна використати для заповнення пустих полів визначеним значенням, яке потім під час аналізу можна використати як *маркер відсутньої інформації*.

Таблиця Excel – це набір рядків, у якому існують більш чіткі, в плані визначень форматів, обмеження, на відміну таблиці Word, у якій не враховується формат даних [13]. Для побудови приймача даних, здатного коректно сприймати всі команди та формати Excel з обробки списків, необхідно керуватися певними принципами та правилами, а саме:

створюваний список повинен мати постійну кількість стовпців;

кількість рядків є змінною, що дає змогу пізніше додати, видалити або переставити записи;

у кожному стовпчику має міститися однотипна інформація;

порожні рядки/записи мають бути видалені, а стовпці заповнені.

*Перша фаза* – витягування. На цій стадії виконується витягування даних із джерела і завантаження їх у допоміжну область (рис. 4). Необхідно проаналізувати та підготувати кінцеву форму приймача даних за принципами описаними вище.

Дані, які отримуємо з джерела, наприклад, мають такий вигляд у форматах Word чи Excel (рис. 5).

У наведеному прикладі таблиці Excel колонки 3 та 14 під час перетворення будуть розділені автоматично на три записи, водночас ці дані стосуються одного запису, тому і в проміжній таблиці мають оброблятися як один запис (запис з одним ідентифікатором). Далі необхідно визначити ключове поле та створити копію форми приймача даних з додатковим (технічним) ключовим полем.

2.	м-р запасу	ІН Олександр Вікторович	15	1 (він- 1963р.н.) розлучений
3.	майор запасу	КІЙ Віктор Володимирович	16	1 (він- 1958р.н.) розлучений

№ п/п	Прізвище, ім'я, по-батькові	Служба	Дата народження	№ документа	Дата запису в список	Категорія	Положення	Служба	Служба	Служба	Служба	Служба	Служба	Служба	Служба
3	Володимир Миколайович		17.12.1997	1											
14	Віктор Володимирович		18.03.2001	2											

Рис. 5. Приклад можливого запису у списку Word та Excel

Якщо позначити кінцеву множину записів, по аналогії з “чорним ящиком”, як множину  $M_{вих}(a, b, c)$ , де  $a, b, c$  – упорядкований набір записів (структуровані дані/записи), тоді джерело даних до перетворення можна зобразити як  $M_{вх}(c, b, a)$ , де  $c, b, a$  – невпорядкований набір записів (неструктуровані дані/записи). Для переходу від однієї множини до іншої необхідний заздалегідь підготовлений і відомий набір інформації (шаблон). Перехід від джерела до приймача математично можна відобразити так:

$$M_{вх}(c, b, a) \Rightarrow M_{дон}(k, c) \cdot M_{дон}(k, b) \cdot M_{дон}(k, a) \Rightarrow M_{вх}(a, b, c)$$

де  $M_{дон}$  – допоміжна множина даних;

$k$  – додаткове (технічне) ключове поле в проміжній таблиці.

Додаткове ключове поле дає змогу врахувати у тимчасових проміжних таблицях відсутність даних (ключове поле = 0), таким чином автоматично визначаючи умову, що ключове поле, за яким визначається однозначність усього запису, має заздалегідь задане відоме і неповторюване значення, а саме значення  $k_i \neq 0$  та  $k_i \neq k_j$ .

Визначившись з кількістю записів та значенням ключового поля, можна переходити до подальшої обробки первинних даних.

Друга фаза перетворення – виявлення помилок. У наведеному прикладі форматі Word (див. рис. 5), у другому стовпчику у записах 2 та 3 міститься за змістом ідентична інформація, але за текстом вона відрізняється, отже має бути приведена до одного вигляду. Подальшу обробку доцільно проводити за стовпчиками, щоб оптимально задіяти можливості Word. Після обробки тексту і первинного розділення даних по полях,

результат аналізується за стовпчиками з використанням вбудованих інструментів сортування та фільтрації [13, 14]. Ці інструменти доцільно використовувати як у Word, під час попередньої обробки, так і в Excel на кінцевих етапах. Вбудований алгоритм сортування значно прискорює процес обробки інформації, що довів ще в XIX ст. американець Герман Холлерит, коли реалізував машинне сортування “окремо для кожного стовбця” [15].

Через інструмент фільтрації проводиться відсіювання “правильних” даних від “шуму”. Фільтрація на етапі первинної обробки даних знижує гостроту проблеми недостовірності, хоча уникнути усіх помилок неможливо. Результатом роботи на цьому етапі буде таблиця (див. рис. 3), в полях якої знаходиться тільки одне значення.

Отже основні етапи другої фази “виявлення помилок” можна визначити так:

список із потрібною інформацією у складі з технічним ключовим полем і колонкою у якій проводитимуться зміни, із формату Excel копіюються у формат Word;

у Word табличний варіант через інструмент перетворення таблиць, перетворюємо на текст, до того ж як роздільник слід виставляти “знак табуляції”;

заміну пропусків між словами доцільно проводити на “знак табуляції”, і далі використати його як роздільник під час формування нової таблиці;

отриманий варіант тексту (у форматі – “слово, знак табуляції, слово”) перетворюємо у табличний варіант, число стовпців запропонує Word. Отримана таблиця копіюється в Excel;

аналіз отриманого результату слід проводити за стовпчиками за допомогою інструменту сортування за низхідним/висхідним порядком.

У разі необхідності усі етапи фази “виявлення помилок” повторюються до досягнення необхідного результату – коли поле містить тільки один елемент даних.

*Третя фаза* – перетворення. Кінцевим результатом має бути текстовий файл з розділювачем, який готовий до завантаження у базу даних. Перетворення, так само, як і виявлення помилок доцільно проводити з використанням технічного (допоміжного) ключового поля. Рекомендації щодо кількості стовпців і використання ключового поля аналогічні попередньому процесу.

Під час запропонованого процесу перетворення даних мають використовуватись загальні правила:

1. Під час формування таблиці у Word, як роздільник доцільно використовувати знак табуляції.

2. Проміжні дані, під час перетворення, необхідно переносити разом з ключовими полями, а у разі відсутності таких ключових полів, необхідно їх створити.

3. Для масового видалення непотрібних даних, доцільно створити шаблон видалення з урахуванням ключового поля.

4. Стандартні ключові поля, для проміжного перетворення доцільно доповнити додатковими цифровими ключовими полями оптимальної довжини.

5. Для проміжного перетворення, таблиці з великою кількістю колонок доцільно розбити на декілька таблиць з максимальною кількістю колонок до шести.

Слід зазначити, що наведений підхід було апробовано під час впровадження підсистем управління нерухомим військовим майном та забезпечення житлом військовослужбовців ІС управління адміністративно-господарськими процесами ЗС України, а також створення макету автоматизованого робочого місця керівника вищого рівня за напрямком управління нерухомим військовим майном. До того ж час на підготовку даних до первинного завантаження із залученням до трьох фахівців було скорочено із запланованого одного року до одного місяця.

Отже на практиці доведено, що дотримання запропонованих правил дає змогу не тільки спростити та оптимізувати первинну підготовку даних для наповнення реляційної бази даних, а також, спростити тестування ІС на етапі її створення. Це також позбавляє розробників можливості зробити “власні помилки”, дає змогу поєднати технологічні прийоми з програмами незалежних

розробників, які виконують спеціалізовані функції, унікальні для конкретного випадку.

**Висновок.** Процес підготовки тестових даних є складним, тривалим і циклічним, тому його оптимізація дасть змогу більш якісно та ефективно використовувати людський і фінансовий ресурс, а під час наповнення бази даних, частково автоматизувати цей процес та звести виникнення помилок, які залежать від людини, до мінімуму.

У статті розглянуто підхід щодо автоматизації процесу підготовки та обробки даних до первинного завантаження в базу даних. У результаті запропоновано послідовність дій на основі використання загальнонавчального інструментарію, проведено формалізацію процесу підготовки даних до первинного завантаження для проведення тестування інформаційних систем.

Напрямом подальших наукових досліджень може бути вирішення питань інтеграції баз даних під час перетворення даних, їх передавання і обробки в інформаційних системах.

#### СПИСОК ВИКОРСТАНОЇ ЛІТЕРАТУРИ

1. Діджиталізація як основний фактор розвитку бізнесу. URL: <http://confmanagement.kpi.ua/proc/article/view/201186> (дата звернення: 02.08.2020).
2. План дій щодо впровадження оборонної реформи. URL: <https://www.mil.gov.ua/diyalnist/reformi-ta-planuvannya-u-sferi-oboroni/22082016-04.html> (дата звернення: 06.08.2020).
3. Концепція інформатизації Міністерства оборони України : затв. наказом Міністерства оборони України від 17.09.2014 р. № 650.
4. Пріоритетні напрямки реформи Міністерства оборони та Збройних Сил України на 2019–2020 роки. URL: <https://www.mil.gov.ua/prioritetni-napryamki-reformi-ministerstva-oboroni-ta-zbrojnih-sil-ukraini-na-2019-2020-roki.html> (дата звернення: 03.08.2020).
5. ERP-системи. URL: <http://techasu.ru/techasu/articles/160-articles2.html> (дата звернення: 02.08.2020).
6. Основи інформаційних систем : навч. посіб. / за ред. проф. В. Ф. Ситника. Київ : КНЕУ, 2007. 252 с.
7. Комп’ютерний контент-аналіз: проблеми та перспективи вирішення. *Методологія, теорія та практика соціологічного аналізу сучасного суспільства* : зб. наук. праць. Вип. 15. Харків : ХНУ, 2009. С. 335–340.
8. IT-платформа онтологічних інформаційно-аналітичних експертних систем. URL: <http://ontology4.inhost.com.ua/?fname=TODOS&detail=ТОДОС> (дата звернення: 02.08.2020).

9. Живицкая Е. Н., Едемская О. П. Системный анализ и проектирование информационных систем : учебно-метод. пособ. для студентов специальности I-40 01 02-02 Информ. системы и технологии в экономике. Минск : БГУИР, 2005. 59 с.
10. Этапы проектирования данных. URL: [http://www.mstu.edu.ru/study/materials/zelenkov/ch\\_5\\_1.html](http://www.mstu.edu.ru/study/materials/zelenkov/ch_5_1.html) (дата звернення: 03.08.2020).
11. Проектирование баз данных. URL: [http://ru.wikipedia.org/wiki/Проектирование\\_баз\\_данных](http://ru.wikipedia.org/wiki/Проектирование_баз_данных) (дата звернення: 04.08.2020).
12. База данных. URL: [http://ru.wikipedia.org/wiki/База\\_данных](http://ru.wikipedia.org/wiki/База_данных) (дата звернення: 02.08.2020).
13. Office: довідка та інструкції. URL: <http://office.microsoft.com/uk-ua/support/#> (дата звернення: 04.08.2020).
14. Устранение неполадок и решение проблем в Microsoft Office XP. URL: <http://support.microsoft.com/ph/2533/uk-ua> (дата звернення: 02.08.2020).
15. Холлерит Г. Галерея портретов. URL: [http://informat444.narod.ru/museum/1\\_17\\_115.htm](http://informat444.narod.ru/museum/1_17_115.htm) (дата звернення: 06.08.2020).

Стаття надійшла до редакційної колегії 24.09.2020

## The justification of approach concerning of data preparation for initial loading for testing Defense Resource Management Information Systems

### Annotation

The implementation of the Unified Defense Resources Management Information System for requires a prudent approach in design of not only the infrastructure, but also unified data model that will be stored in a unified repository.

Creation of such model requires a properly design data, which in turn requires proper formation and its preparation. This process is quite long, complex and cyclical and will include certain stages of transformation, extraction and structuring of information.

*The main purpose of the article* is to simplify the process of preparing primary data for loading into a repository with relational connections. Make this process repetitive and as independent as possible from person. Using this approach, we can minimize or even avoid the use of expensive software platforms and not involve third-party professionals. Such optimization will allow more efficient and rational use of software and human resources partially automate it and, if necessary, parallel the process and minimize errors.

The article considers a modern approach how to automate the process of preparation and processing of data for initial uploading data into the database. As a result, a sequence of actions based on the use of commonly used set of tools is proposed, the process of data preparation for the initial download for testing information systems is formalized.

The approach is given was tested during the introduction of subsystems for the management of military real estate and the provision of housing for servicemen of the Information System Management of administrative and economic processes of the Armed Forces of Ukraine. All at once, the time for preparing data for the initial download with the involvement of up to three specialists was reduced from the planned one year to one month.

**Keywords:** defense resource management; preparation of data for testing; data conversion; data structuring; structured data.