

ГРАМАТИЧНА МОДЕЛЬ ПЕРЕТВОРЕННЯ EDIFACT-ПОВІДОМЛЕНЬ НА XML-ДОКУМЕНТИ

На основі описаного у стандарті ISO 9735 набору нотацій синтаксису EDIFACT-мови побудовано контекстно-вільну граматику для перевірки EDIFACT-повідомлень. Розроблено механізм перетворення дерева граматичного розбору EDIFACT-повідомлення на відповідні XML-схему і XML-повідомлення, погоджений з описаними у стандарті ISO/TS 20625:2002 правилами генерації XML-схем.

Вступ. Економічні переваги електронного обміну даними (Electronic Data Interchange, EDI) загальноновизнано, однак витрати на встановлення EDI-зв'язку все ще завеликі через докладне узгодження потреб бізнес-партнерів. Початкові значні витрати на досягнення таких угод не виправдовує короткострокова співпраця, також мало ефективні реалізації з численними партнерами і видами угод. Сьогодні EDI-системи досить дорогі (від 10 до 100 тис. дол США) і недоступні багатьом дрібним користувачам. Більшість EDI-систем рентабельні лише у довгостроковій співпраці обмеженого числа партнерів.

Подолати ці бар'єри вдалося завдяки стандартизації бізнес-сценаріїв і послуг підтримки EDI-систем. Якщо бізнес-сценарій узгоджено, а реалізації відповідають стандартам Open-EDI, то немає потреби у попередньому узгодженні дій бізнес-партнерів, окрім рішення брати участь у Open-EDI-транзакції згідно з бізнес-сценарієм. Сьогодні в EDI-системах задіяно 12 стандартів Open-EDI, які визначають формати повідомлень і забезпечують перетворення між різними форматами і протоколами обміну. У 1996 р. для синхронізації стандартів ООН установила правила UN/EDIFACT-обміну електронними даними для адміністрації, торгівлі й транспорту, зафіксовані у 10-томно-

му стандарті ISO 9735, що пережив уже три покоління [2-12]. В ISO/IEC 14662:1997 введено еталонну Open-EDI-модель [1], створену для електронної обробки бізнес-транзакцій самостійних корпорацій зі складною структурою зв'язків між підрозділами (наприклад, загальні, особисті, виробничі, територіальні зв'язки).

Еталонна Open-EDI-модель відбиває два погляди на опис суттєвих аспектів бізнес-транзакцій (див. рис. 1). Бізнес-операційний погляд BOV націлено на семантику даних і правила бізнес-транзакцій (операційні угоди, взаємні зобов'язання щодо бізнес-вимог Open-EDI), таких як у стандарті ISO 9735. Функціонально-сервісний погляд FSV відбиває підтримку послуг інформаційними технологіями (інтерфейси і протоколи, механізми перекладу і захисту тощо), наприклад, як в ISO/TS 20625:2002.

Нині проходить затвердження розроблений проект Національного стандарту України, гармонізованого з ISO 9735:2002. У ньому визначено загальні правила синтаксису для форматування пакетних та інтерактивних EDIFACT-повідомлень для обміну EDI-систем, а також службові каталоги нотацій синтаксису, що містять понад 200 типів службових сегментів повідомлень.

Якщо бізнес-партнери бажають синхронізувати власні внутрішні EDI-системи із зовніш-

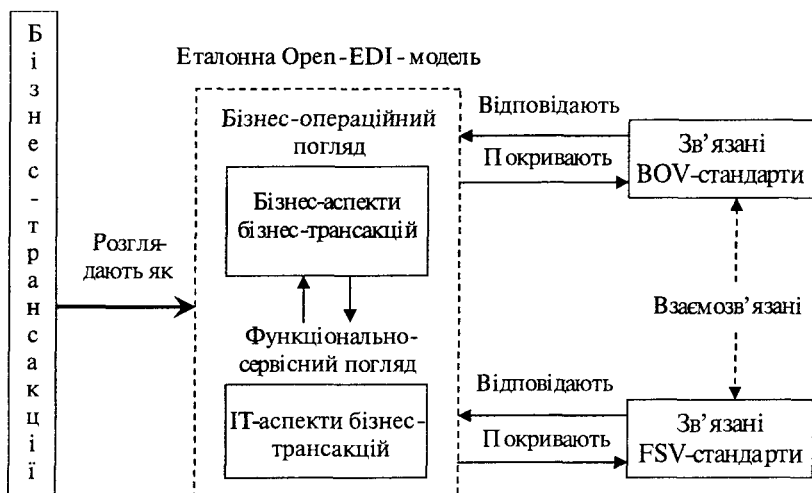


Рис. 1. Open-EDI-середовище

нши системами своїх партнерів, виникає потреба у побудові спеціального сучасного інтерфейсу на основі мережних сервісів, насамперед мовою XML. Інтерес до XML/EDI-систем виявляють і в Україні. У цій статті розглянуто граматичну модель для побудови TaKOїXML/EDI-системи, узгодженої з положеннями еталонної Open-EDI-моделі.

Системи XML/EDI та їх компоненти. Уводячи XML/EDI-системи, що містять програмні агенти і глобальний словник, навіть маленька компанія може легко інтегруватися у бізнес-процеси. Девід Веббер [2] провів порівняння трьох технологій: (1) традиційних EDI-систем, (2) електронної комерції, що використовує HTML-форми і Web-сайти, та (3)XML/EDI-рішень. Проаналізовано фактори, суттєві для електронного підприємства (electronic enterprise):

- низька вартість транзакцій;
- інтеграція з існуючими бізнес-системами;
- знижені продажі й цикли обміну інформацією;
- глобальна досяжність;
- функціональність оброблення вмісту як тексту;
- ефективне оброблення великої кількості бізнес-партнерів;
- простота обслуговування і розгортання;
- універсально доступні програмні інструменти;

- зворотна сумісність;
- масштабованість з майбутніми вимогами і технологіями;
- наявність кращих бізнес-методів;
- точність, безпека, надійність і відмовостійкість;
- обслуговування у середовищі з великим обсягом транзакцій.

Аналіз проведено за 10-бальною шкалою (0 — повна відсутність фактору; 5 — фактор використано на середньому рівні; 10 — фактор використано повною мірою). Як видно з табл. 1, HTML на 50 %, а XML/EDI на 100 % краще за традиційні EDI-системи.

Мета XML/EDI — установити відкриті й загальнодоступні стандарти для комерційного обміну даними, забезпечивши засоби для задоволення будь-яких бізнес-потреб. Для цього необхідна методологія, широко і вільно розповсюджена, легко адаптована і розширювана під майбутні вимоги і нові технології. Мова XML забезпечує таку доступну і мобільну методологію для контрольованого обміну даними, оскільки загалом призначена для обміну інформацією у формі комп'ютерних візуалізованих «документів». Але не всі комерційні дані передають у візуалізованій формі. Наприклад, специфічні дані EDIFACT-повідомлень перед відображенням на екрані потрібно перетворити за правилами оброблення, використовуючи певні шаблони.

Таблиця 1. Порівняння технологій EDI, XML/EDI і Web-технологій

Метод	A	B	C	D	E	F	G	H	I	J	K	L	M	Всього
EDI	5	3	5	3	0	3	3	0	8	3	5	7	10	55
HTML	8	3	7	8	5	3	5	8	3	8	7	5	5	75
XML/EDI	9	10	9	8	9	10	8	8	10	9	10	9	8	117

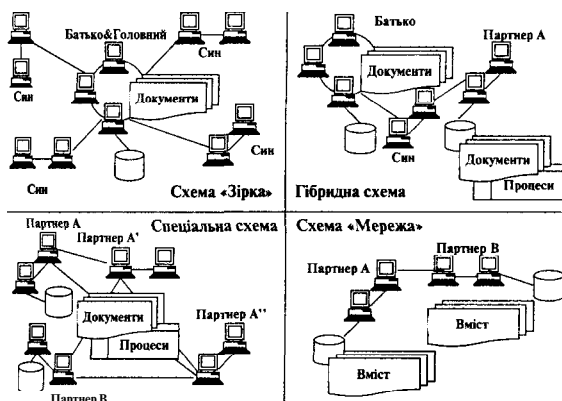


Рис. 2. Чотири основні моделі XML/EDI

Відомі чотири динамічні моделі обміну XML/EDI (див. рис. 2), що включають традиційні методи розгортання EDI поряд з новими документними можливостями XML. У класичній EDI-схемі «Зірка» головний бізнес-партнер установлює стандарти для своїх бізнес-партнерів. Спеціальна схема (ad hoc) — це нова модель, основана на мережі. Менші бізнес-партнери взаємодіють за спеціальними зв'язками, за потреби встановлюючи формальніші методи. Гібридна схема — це комбінація перших двох моделей. У ній бізнес-партнери розширюють схему «Зірка», створюючи нові версії структур і зв'язуючи з ними свої спеціальні підрозділи. У документній моделі «Мережа» вміст як найцінніші дані обміну можна отримати за заздалегідь заданими правилами, запитати або переслати, як у класичному прикладі електронного каталогу.

Раніше в EDI-обміні від одного комп'ютера до іншого переміщували поля даних без втручання людини. Нова XML/EDI-модель переважно заснована на інтерактивному виборі набору опцій і заповненні електронних форм за вимогами користувача. Внаслідок цього фундаментально змінився спосіб оброблення даних, для якого ланцюжок обробки інформації «створення-жердача-юдержання» замінено на концепцію активних об'єктів, яким притаманні процеси, основані на класі інформації, що міститься в них. Сучасний рахунок-фактуроване більше не містить копію інформації, збереженої в базі даних, з якої його згенеровано; замість цього він містить покажчик, звідкіля можна одержати дані, і їх витягають з керованого джерела під час кожного оброблення рахунка-фактури.

XML/EDI — це щось більше за пряме перенесення XML-документів у EDI-систему. У рамках XML/EDI інтегровано п'ять технологій

(див. рис. 3); кожен компонент додає свої специфічні можливості. Розглянемо докладніше кожен зі складових частин.

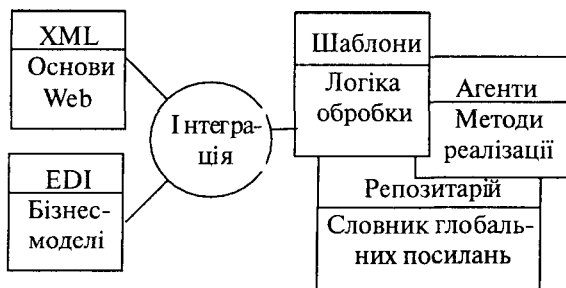


Рис. 3. Складові частини XML/EDI технологій

XML. За допомогою заздалегідь визначених тегів визначено об'єктну модель даних, яка згодом заповнюється даними і передається як електронний документ. Ідентифікатори EDI-сегментів замінюють на теги XML, або частина даних з EDI-сегмента додається у теги як параметри.

EDI. Розроблені в EDI-системах стандарти здатні подавати дані в простому форматі, який однозначно інтерпретують сторони — приймач і передавач. XML/EDI забезпечує стовідсоткову сумісність з EDI-системами, використовуючи обмін EDIFACT-повідомленнями. Розробка протоколів XML/EDI дає змогу застосовувати існуючі EDI-системи і не вимагає нових капіталовкладень.

Шаблони (Templates) — це набір правил для керування обробленням бізнес-даних на клієнтській і серверній стороні. Шаблон можна завантажити з віддаленого джерела, звідки прийшов XML-документ, у складі якого може бути шаблон. Шаблони використовують Document Type Definitions (DTD's), за якими визначено об'єктну модель даних. Віддалене використання (DTD's) надасть можливість усім клієнтським застосуванням однозначно описувати використовувану модель даних.

Агенти інтерпретують шаблони, щоб інтерактивно виконати транзакції і взаємодіяти з користувачем. Розбирання структури XML-агент може здійснювати прямо на комп'ютері клієнта і використовувати при цьому потрібні для користувача дані та їх подання. Спочатку агентами керували шаблони, але вже розроблено окремі протоколи для агентів.

Репозитарій — спільні загальнодоступні в Інтернеті словники, які вже використовують традиційні EDI-системи, щоб знайти значення й область визначення EDI-елементів. Спільно використовувани загальні словники забезпечу-

ють автоматичні пошукові таблиці більш гнучким і сучасним механізмом пошуку як семантичною основою EDI-транзакцій.

Нині міжнародна спільнота відпрацьовує такі інтерфейси рівнів XML/EDI:

- стандартні транспортні механізми Інтернет-обміну і збереження файлів;
- формати подання даних і передачі повідомлень;
- синтаксис XML-даних;
- правила граматичного розбору XML-документа і створення об'єктної моделі;
- правила XSL-подання і зв'язування зі сценаріями і рівнями розбору;
- правила керування даними користувачьких застосувань та інтерфейсів.

Сьогодні доступні синтаксичні XML-аналізатори (парсери), програми перегляду XML-документів (браузери), програми розмітки сторінок і бібліотеки бізнес-об'єктів. За допомогою Web-браузера можна взаємодіяти з будь-якими компонентами системи, використовуючи XML-подання і XQL-запити. Передбачається вбудову-

вати XML/EDI-компоненти віснуючі програмні продукти для значного прискорення процесу створення нових застосувань.

Первинні компоненти XML/EDI подають загальну мову опису і синтаксичні правила та включають таблицю визначення даних (DTD); репозитарій (сховище) даних; сегменти й елементи даних (тобто EDIFACT, X12 або BSI-каталоги); бібліотеки бізнес-об'єктів; Web-сайти бізнес-партнерів з пов'язаними корпоративними базами даних [13].

Нотації синтаксису EDIFACT-мови. В ISO 9735 обмін у EDI-системах визначено як «послідовність повідомлень і/чи пакетів одного або різних типів, які починає заголовок обміну (чи службовий рядок повідомлення) і закінчує завершувач обміну». Обмін має включати лише:

- повідомлення;
- пакети;
- повідомлення і пакети;
- групи, що містять повідомлення;
- групи, що містять пакети;

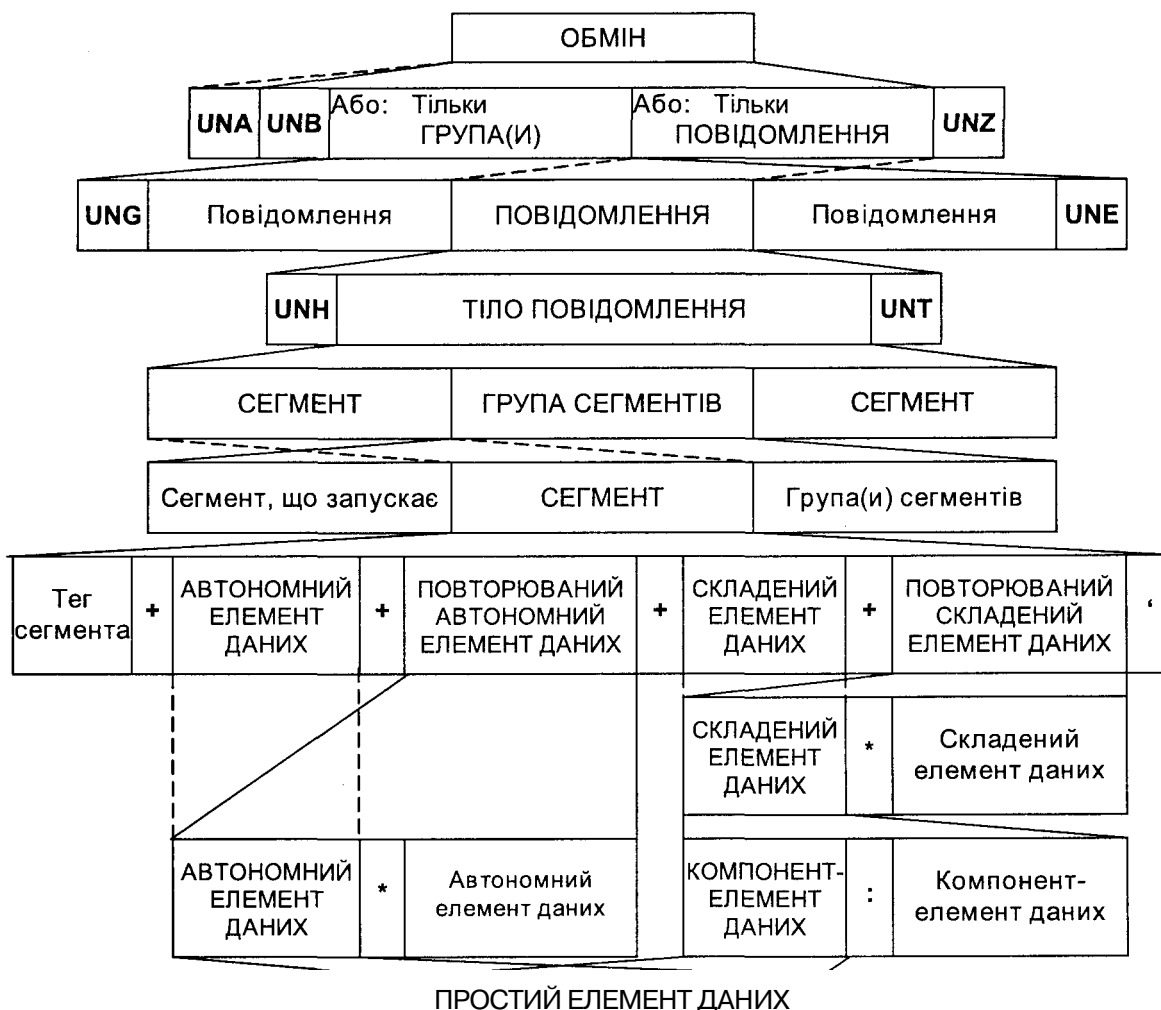


Рис. 4. EDI-повідомлення в транзакції

- групи, що містять повідомлення і пакети.

Стандартизовано структури таких EDIFACT-категорій, як групи, повідомлення, групи сегментів, сегментів, пакетів, простого і складеного елементу даних.

Група включає заголовок групи, повідомлення(нь) і завершувач (кінцевик) групи. Повідомлення містить заголовок, тіло і завершувач (кінцевик). Тіло повідомлення складається із сегмента(ів) і/чи групи(груп) сегментів, а група сегментів – з пускового сегмента, сегмента(ів) і, можливо, групи(груп) сегментів. Сегмент включає тег сегмента й автономний(ні) елемент(и), і/чи складений(ні) елемент(и), і/чи повторюваний(ні) автономний(ні) елемент(и), і/чи повторюваний(ні) складений(ні) елемент(и) даних.

Повторюваний автономний елемент – це одна чи більше появ того самого автономного елемента даних. Повторюваний складений елемент – це одна чи більше появ того самого складеного елемента даних. Складений елемент – це два чи більше компонент-елементи даних. Компонент-елемент – це простий елемент даних. Автономний елемент даних – це простий елемент даних. Простий елемент даних включає одне значення елемента даних.

Графічне подання EDIFACT-повідомлення з EDI-обміну показано на рис. 4.

Якщо повідомлення готове до передачі, до нього застосовують визначені в ISO 9735 правила для включення, вилучення і блокування/придушення символів у елементах даних. Правила для включення і вилучення – це визначення наявності, включення груп сегментів, вилучення груп сегментів, включення/вилучення сегментів, включення/вилучення елементів даних. Правила блокування використовують, щоб придушувати (не включати в передачу) такі символи, як незначущі символи, значущі нулі й пробіли.

Граматика EDIFACT-мови. На основі описаної в ISO 9735 сукупності нотацій синтаксису для формату EDIFACT-повідомлення граматику задамо у вигляді бекусовських нормальних форм (БНФ):

EDI_G = {EDI_T, EDI_N, EDI_P, <ОБМІН>},
де EDI_T – алфавіт термінальних символів EDIFACT-мови, які входять до продукцій в одинарних лапках ‘ (наприклад, ‘UNZ’ чи ‘UNA’); EDI_N – алфавіт нетермінальних символів, тобто множина металінгвістичних змінних у кутових дужках; EDI_P – кінцева множина правил-продукцій, пронумерованих від 1 до 64; <ОБМІН> – початковий символ граматики EDI_G; * – опе-

рація конкатенації всередині продукції, ε – пустий символ.

Набір БНФ-продукцій граматики EDIFACT_G має вигляд:

- (1) <ОБМІН> ::= ‘UNA’*‘UNB’* <УМІСТ>*‘UNZ’
- (2) <УМІСТ> ::= <ПОВІДОМЛЕННЯ>
- (3) <УМІСТ> ::= <ПАКЕТИ>
- (4) <УМІСТ> ::= <ПовідомТаПак>
- (5) <УМІСТ> ::= <ГрупаПовідом>
- (6) <УМІСТ> ::= <ГрупаПак>
- (7) <УМІСТ> ::= <ГрупаПовідомТаПак>
- (8) <ПОВІДОМЛЕННЯ> ::= <Повідомлення>*<ПОВІДОМЛЕННЯ>
- (9) <ПОВІДОМЛЕННЯ> ::= <Повідомлення>
- (10) <ПАКЕТИ> ::= <Пакет>*<ПАКЕТИ>
- (11) <ПАКЕТИ> ::= <Пакет>
- (12) <ПовідомТаПак> ::= <ПОВІДОМЛЕННЯ>*<ПАКЕТИ>
- (13) <ГрупаПовідом> ::= ‘UNG’*<ПОВІДОМЛЕННЯ>*‘UNE’*<ГрупаПовідом>
- (14) <ГрупаПовідом> ::= ‘UNG’*<ПОВІДОМЛЕННЯ>*‘UNE’
- (15) <ГрупаПак> ::= ‘UNG’*<ПАКЕТИ>*‘UNE’*<ГрупаПак>
- (16) <ГрупаПак> ::= ‘UNG’*<ПАКЕТИ>*‘UNE’
- (17) <ГрупаПовідомТаПак> ::= ‘UNG’*<ПОВІДОМЛЕННЯ>*<ПАКЕТИ>*‘UNE’*<ГрупаПовідомТаПак>
- (18) <ГрупаПовідомТаПак> ::= ‘UNG’*<ПОВІДОМЛЕННЯ>*<ПАКЕТИ>*‘UNE’
- (19) <Повідомлення> ::= ‘UNH’*<Специф Повідомлення>*<Тіло>*‘UNT’
- (20) <Специф Повідомлення> ::= <ЗаголовокСпециф>*<Опис>*<ЗавершСпециф>
- (21) <Опис> ::= <Позиція>*<Статус>*<КількістьПояв>*<Опис>
- (22) <Опис> ::= <Позиція>*<Статус>*<КількістьПояв>
- (23) <Статус> ::= <Обов’язковий>
- (24) <Статус> ::= <Умовний>
- (25) <Тіло> ::= <Сегменти>*<ГрупиСегментів>
- (26) <Тіло> ::= <Сегменти>
- (27) <Тіло> ::= <ГрупиСегментів>
- (28) <Сегменти> ::= <Сегмент>*<Сегменти>
- (29) <Сегменти> ::= <Сегмент>
- (30) <ГрупиСегментів> ::= <ГрупаСегментів>*<ГрупиСегментів>
- (31) <ГрупиСегментів> ::= <ГрупаСегментів>
- (32) <ГрупаСегментів> ::= <СегментЩоЗапускає>*<Сегменти>

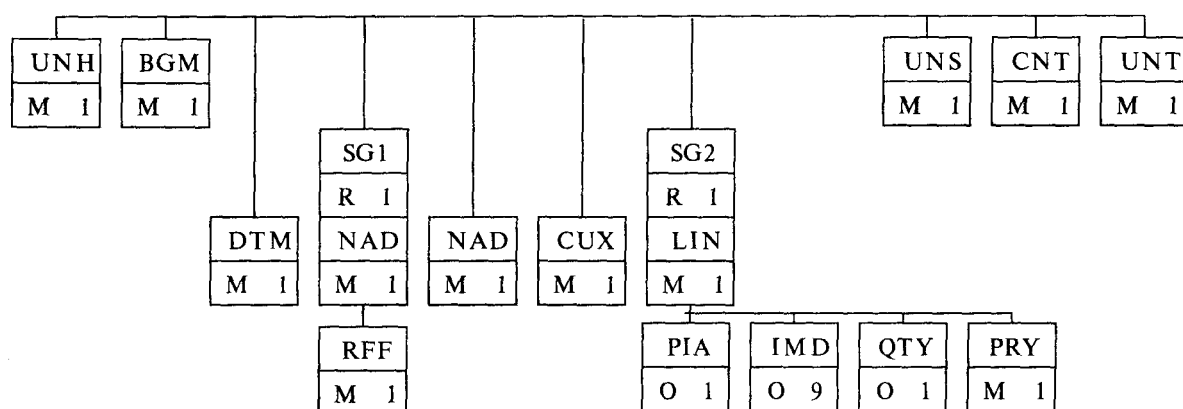


Рис. 5. Структура EDIFACT-повідомлення ORDERS

- (33) <ГрупаСегментів> ::= <ЗалежнаГрупаСегментів>
 (34) <ЗалежнаГрупаСегментів> ::= <Сегмент>*<ЗалежнаГрупаСегментів>
 (35) <ЗалежнаГрупаСегментів> ::= <Сегмент>
 (36) <ЗавершСпециф> ::= '''
 (37) <Сегмент> ::= <ТегСегмента>*<СпецифСегмента>*<Дані>*'''
 (38) <Сегмент> ::= <ТегСегмента>*<СпецифСегмента>*<Дані>
 (39) <Дані> ::= <ЕлементиДаних>
 (40) <Дані> ::= <СкладЕлементи>
 (41) <Дані> ::= <ЕлементиДаних>*<СкладЕлементи>
 (42) <ТегСегмента> ::= <ПростийЕлемент>
 (43) <СкладЕлементи> ::= <СпецифСкладЕлементів>*<КомпЕлемент>*<КомпЕлемент>*<КомпЕлементи>
 (44) <СкладЕлементи> ::= <СпецифСкладЕлементів>*<КомпЕлемент>*<КомпЕлементи>
 (45) <КомпЕлементи> ::= <КомпЕлемент>*<КомпЕлементи>
 (46) <КомпЕлементи> ::= <КомпЕлемент>
 (47) <КомпЕлементи> ::= ε
 (48) <ЕлементиДаних> ::= <ПростийЕлемент>*<ЕлементиДаних>
 (49) <ЕлементиДаних> ::= '*'
 (50) <ЕлементиДаних> ::= <ПростийЕлемент>
 (51) <ЕлементиДаних> ::= ε
 (52) <КомпЕлемент> ::= <ПростийЕлемент>*'.'
 (53) <КомпЕлемент> ::= <ПростийЕлемент>
 (54) <КомпЕлемент> ::= '.'
 (55) <Пакет> ::= <ЗаголовокОб'єкта>*<Об'єкт>*<ЗавершОб'єкта>
 (56) <ПростийЕлемент> ::= <Об'єкт>*'+'
 (57) <ПростийЕлемент> ::= <Об'єкт>

- (58) <ПростийЕлемент> ::= '+'
 (59) <ЗаголовокСпециф> ::= ε
 (60) <Опис> ::= ε
 (61) <ЗавершСпециф> ::= ε
 (62) <Сегменти> ::= ε
 (63) <ГрупиСегментів> ::= ε
 (64) <СпецифСегмента> ::= ε

Оскільки всі продукції записано у вигляді БНФ, граматику є контекстно-вільною. Використовуючи ліве спадне виведення, побудуємо дерево граматичного розбору для EDIFACT-повідомлення ORDERS, яке традиційно використовують у демонстрації механізму перекладу XML/EDI (див., наприклад, [13]). У замовленні ORDERS на придбання дев'яти найменувань товарів повідомлення включає нотації синтаксису, що належать тільки тілу повідомлення. Закінчене EDIFACT-повідомлення, готове до передачі, має включати й іншу, службову інформацію про процедуру обміну.

У другому стовпці табл. 2 наведено текст EDIFACT-повідомлення ORDERS, розбитого на сегменти. Ліворуч квадратними дужками позначено групи сегментів SG1 і SG2. У стовпці R зазначено кількість повторень сегмента, а в стовпці S – статус кожного сегмента: C (Conditional) – умовний, M (Mandatory) – обов'язковий, O (Optional) – необов'язковий, R (Required) – потрібний. На рис. 5 показано структуру EDIFACT-повідомлення ORDERS.

Побудуємо дерево граматичного розбору EDIFACT-повідомлення ORDERS. Це дерево важко показати графічно, хоча його топологія аналогічна структурі, зображеній на рис. 5, покажемо його лінеаризовано чи послідовно. Запис \Rightarrow (59,60,61) \Rightarrow використано між двома пропозиціями в сентенціальній/пропозиційній формі для вказування, що правила 59, 60, 61 застосовано для породження другого речення з першого в лівому виведенні.

Таблиця 2. Опис сегментів і полів EDIFACT-повідомлення ORDERS

№	Сегменти повідомлення	S	R	Вміст сегмента
00	UNA UNB			Заголовок обміну
01	UNH+000002+ORDERS:D:96A:UN:EAN008'	M	1	Заголовок повідомлення
02	BGM+220+B00002+9'	M	1	Номер замовлення
03	DTM+137:19940202:102'	M	1	Дата відправки повідомлення
04	Група сегментів SG1 NAD+BY+++Stadt-und Universitätsbibliothek:Frankfurt+Bockenheimer Landstr. 134-13 8+Frankfurt++ 60325'	R	1	Покупець
05	RFF+API:DE1141110388'	M	1	Ідентифікатор покупця
06	NAD+SU+++DREIER'	M	1	Найменування постачальника
07	CUX+2:EUR:9'	M	1	Валюта оплати – євро
08	Група сегментів SG2 LIN+1'	R	1	
09	PIA+5+3772815359:IB'	M	1	Позиція замовлення 1
10	IMD+F+050+:::Die "Jahrbuecher fuer wissenc haftliche Kritik"	O	1	ISBN-ідентифікатор замовлення
...	...			
20	QTY+21:1'	O	1	Кількість копій замовлення
21	PRI+AAE:295:CA'	M	1	Ціна замовлення
22	UNS+S'	M	1	Розділювальний сегмент
23	CNT+2:1'	M	1	Загальна кількість позицій – 1
24	UNT+25+000002'	M	1	Загальна кількість сегментів – 25
25	UNT	M	1	Завершувач повідомлення
26	UNZ			Завершувач обміну

Запис $\Rightarrow ((28), 37, 39, 48, 57, 52, 58) \Rightarrow$ означає, що перед застосуванням правил 37, 39, 48, 57, 52, 58 зліва направо використано правило 28 для породження додаткового нетермінального символу, наприклад, щоб породити додатковий сегмент. Запис $\Rightarrow (\text{for } (i=1; i \leq 21; i++) \text{ do } \{(28), 37, 39, 48, 57, 52, 58\}) \Rightarrow$ означає, що для одержання другого речення з першого (розташованого в лівій частині чи перед символом '='>) у виведенні застосовано набір правил $\{(28), 37, 39, 48, 57, 52, 58\}$, відтак враховано семантику $\Rightarrow ((28), 37, 39, \dots, 58)$. Набір правил застосовано 21 раз і прийнято на вході $\{(28), 37, 39, \dots, 58\}$ речення сентенціальної/пропозиційної форми, створеної на попередньому кроці.

Побудуємо ліве виведення повідомлення ORDERS.

ОБМІН $\Rightarrow (1) \Rightarrow$ UNA UNB*УМІСТ*UNZ
 $\Rightarrow (2) \Rightarrow$ UNA UNB*ПОВІДОМЛЕННЯ*UNZ
 $\Rightarrow (9) \Rightarrow$ UNA UNB*Повідомлення*UNZ
 $\Rightarrow (19) \Rightarrow$ UNA UNB UNH*СпецифПовідомлення*Тіло*UNT UNZ $\Rightarrow (20) \Rightarrow$ UNA UNB UNH*ЗаголовокСпециф*Опис*ЗавершСпециф*Тіло*UNT UNZ $\Rightarrow (59, 60, 61) \Rightarrow$ UNA UNB UNH*Тіло*UNT*UNZ $\Rightarrow (25) \Rightarrow$ UNA UNB UNH*Сегменти*ГрупиСегментів*UNT UNZ $\Rightarrow (28) \Rightarrow$ UNA UNB UNH*Сегмент*Сегменти*ГрупиСегментів*UNT*UNZ $\Rightarrow (37, 62, 63) \Rightarrow$ UNA UNB UNH* Тер Сегмента* Специф Сегмента* Дані*Сегменти*

UNT UNZ $\Rightarrow (42, 47, 57, 64, 39, 54, 58, 36) \Rightarrow$ UNA UNB UNH+000002+ORDERS:D:96A:UN:EAN008' *Сегменти*'UNT*'UNZ' $\Rightarrow (28) \Rightarrow$ UNA UNB UNH+000002+ORDERS:D:96A:UN:EAN008' *Сегмент*Сегменти*UNT UNZ $\Rightarrow (37, 39, 48, 57, 52, 58,) \Rightarrow$ UNA UNB UNH+000002+ORDERS:D:96A:UN: EAN008' BGM+220+B00002+9'*Сегменти* UNT UNZ $\Rightarrow (28) \Rightarrow$ UNA UNB UNH + 000002+ORDERS:D:96A:UN:EAN008' BGM+220+ B00002+9' Сегмент*Сегменти* UNT UNZ $\Rightarrow ((28), 37, 39, 48, 57, 52, 58) \Rightarrow$ UNA UNB UNH+000002+ORDERS:D:96A:UN: EAN008' BGM+220+B00002+9'DTM+137:19940202:102'* Сегменти* UNT UNZ $\Rightarrow ((28), 37, 39, 48, 57, 52, 58) \Rightarrow$ UNA UNB UNH+000002+ORDERS: D:96A: UN:EAN008' BGM+220+B00002+9'DTM+137:19940202:102'NAD+BY+++Stadt-undUniversitaetsbibliothek*Сегменти* UNTUNZ $\Rightarrow ((28), 37, 39, 48, 57, 52, 58) \Rightarrow$ UNA UNB UNH+000002+ORDERS:D:96A: UN: EAN008' BGM+220+B00002+9'DTM+137:19940202:102'NAD+BY+++Stadt-und Universitaetsbibliothek: Frankfurt+Bockenheimer Landstr.134-138+ Frankfurt++60325'RFF+API: DE1141110388'*Сегменти*UNTUNZ $\Rightarrow (\text{for } (i=1; i \leq 21; i++) \text{ do } \{(28), 37, 39, 48, 57, 52, 58\}) \Rightarrow$ UNA UNB UNH+000002+ORDERS:D:96A:UN: EAN008' BGM+220+B00002+9' DTM+137:19940202:102'NAD+BY+++StadtundUniversitaetsbibliothek: Frankfurt

+Bockenheimer Landstr. 134-13 8+ Frankfurt++
60325'RFF+API:DE 1141110388' NAD+SU
+++DREIER' CUX+2:DEM:9' LIN+1' PIA+
5+3772815359: IB' IMD+F+050+:::Die
"Jahrbuecher fuer wissensc haftl: iche Kritik"
IMD+F+060+:::Hegels Berliner Gegenakademie'
IMD+F+065+:::Hrsg.von Christoph Jamme'
IMD+F+110+:::Stuttgart-Bad Cannstadt'
IMD+F+120+:::Frommann-Holz-boog' IMD+
F+170+ :::1994'IMD+F+190+:::Spekulation und
Erfahrung'IMD+F+191+:::Abt.2'IMD+F+192
+::: Untersuchungen'QTY+21:1'PRI+AAE:295:
CA' UNS+S'CNT+2:1'CNT+2:1'UNT+25+
000002'UNT UNZ

Оскільки побудову виведення завершено успішно, то EDIFACT-повідомлення ORDERS сформовано правильно, що дає підставу сформулювати спочатку XML-схему, а потім XML-документ.

Переклад EDIFACT-повідомлення на XML-документ. В ISO/TS 20625:2002 [12] визначено правила одержання XML-схем з UN/EDIFACT і метод подання семантичних даних. У ньому описано правила формування імен тегів, перетворення структури EDI-повідомлення в XML-схеми, оптимізації XML-схем, перетворення форматів елементів даних та інші аспекти. Згідно з ISO/TS 20625:2002 у формуванні імен XML-тегів компонентам EDIFACT-повідомлення залежно від їхнього рівня в структурі повідомлення присвоюють відповідний префікс і необов'язковий суфікс, щоб задати семантику EDI-елементів. Наприклад, для EDIFACT-повідомлення ORDERS використовуємо такі імена тегів:

"M_" +тип повідомлення
+[суфікс] M_ORDERS
"G_" +група сегментів+[суфікс] G_SG1
"S_" +сегмент
+[суфікс] S_UNH, S_NAD_seller,
S_NAD_buyer

"C_" +складений елемент даних
+[суфікс] C_C507
"D_" +елемент даних+[суфікс] D_2005

Кожному EDI-компоненту (типу повідомлення, групі сегментів, сегменту, елементу даних тощо) відповідає окремий XML-елемент, згенерований за деревом граматики розбору (див., наприклад, рис. 5).

Повідомлення ORDERS містить два сегменти NAD; один задає ім'я й адресу покупця в групі сегментів SG1, а другий – найменування постачальника в тілі повідомлення M_ORDERS. У XML-схемі кожному сегменту NAD відповідає окремий тег, до імені якого додано суфікс, що відбиває семантику: S_NAD_buyer і S_NAD_seller. Простір імен 'din' використано у прикладі для ілюстрації, його можна опустити або замінити на інший зручний простір імен. Згідно з ISO/TS 20625:2002 статус XML-елементів задає атрибут minOccurs (мінімальна кількість появ): EDI-статусу «обов'язковий» (M=Mandatory) відповідає значення атрибута "1" статусу «умовний» (C=Conditional) – значення атрибута "0". Структуру EDIFACT-повідомлення ORDERS перетворимо на XML-схему такого вигляду:

```
<xsd:element name="M_ORDERS">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element ref="din:S_UNH"/>
      <xsd:elementref="din:S_BGM"minOccurs=
"1" maxOccurs="1"/>
      <xsd:element ref="din:S_DTM" minOccurs=
"1" maxOccurs="1"/>
      <!--Група сегментів NAD_buyer і RFF-->
      <xsd:element ref="din:S_NAD_seller"
minOccurs="1" maxOccurs="1"/>
      <xsd:element ref="din:S_CUX" minOccurs=
"1" maxOccurs="1"/>
      <!--Група сегментів LIN, PIA, IMD, QTY і
PRI-->
```

Таблиця 3. Опис сегмента DTM у довіднику сегментів EDSD

Тег	Назва елементів даних сегмента DTM		S	Формат	Вміст
C507	DATA/TIME/PERIOD	ДАТА/ЧАС/ПЕРІОД	M		
2005	Data/time/period qualifier	Специфікатор да- ти/часу/періоду	M	an..3	+137
2380	Data/time/period	Дата/час/період	C	an..35	:19940202
2379	Data/time/period format qualifier	Специфікатор формату да- ти/часу/періоду	C	an..3	102'


```

<xsd:element ref="din:S_SG2" minOccurs="1"
maxOccurs="1"/>
<xsd:element ref="din:S_UNH" minOccurs="1"
maxOccurs="1"/>
<xsd:element ref="din:S_CNT" minOccurs="1"
maxOccurs="1"/>
<xsd:element ref="din:S_UNT" minOccurs="1"
maxOccurs="1"/>
</xsd:sequence>
<xsd:attribute name="EDIPath" type="xsd:string"
fixed="ORDERS"/>

```

!-Атрибут EDIPath містить посилання на вихідний стандарт EDI-->

```

</xsd:complexType>
</xsd:element>

```

Згідно з довідником сегментів EDSD з низки стандартів UN/EDIFACT кожен сегмент містить групу елементів даних. Так, у табл. 3 наведено опис сегмента дати DTM.

У правому стовпці табл. 3 для ілюстрації зазначено вміст елементів даних сегмента DTM повідомлення ORDERS: DTM+137: 19940202: 102'. Відповідна цьому опису XML-схема матиме вигляд:

```

<xsd:element name="S_DTM">
<xsd:complexType>
<xsd:sequence>
<xsd:element ref="din:D_2005"/>
<xsd:element ref="din:D_2380"/>
<xsd:element ref="din:D_2379"/>
</xsd:sequence>
<xsd:attribute name="EDIPath" type="xsd:
string" fixed="ORDERS.DTM(0030)"/>
</xsd:complexType>
</xsd:element>
...
<!-- Елемент даних D_2005 сегмента DTM-->
<xsd:element name="D_2005">
<xsd:complexType>
<xsd:simpleContent>
<xsd:extension base="din:D_2005">
<xsd:attribute name="EDIPath" type="xsd:
string" fixed=" ORDERS.DTM. C507.2005
(0030:010:01)"/>
</xsd:extension>
</xsd:simpleContent>
</xsd:complexType>
</xsd:element>

```

!-Елементи даних D_2380 і D_2379 сегмента DTM-->

...

Приклад XML-повідомлення для наведених вище фрагментів схеми:

...

```

<din:M_ORDERS xmlns:din="..."xmlns:xsi="
..." xmlns:schemaLocation="...">

```

```

<!--Сегменти S_UNH, S_BGM і S_DTM-->
<din:S_UNH>

```

...

```

</din:S_UNH>

```

```

<din:S_BGM>

```

...

```

</din:S_BGM>

```

```

<din:S_DTM>

```

```

<din:D_2005>137</din:D_2005>

```

```

<din:D_2380>19940202</din:D_2380>

```

```

<din:D_2379>102</din:D_2379>

```

```

</din:S_DTM>

```

```

<!--Група сегментів G_SG1-->

```

```

<din:G_SG1>

```

...

```

</din:G_SG1>

```

```

<!--Сегменти S_NAD_seller і S_CUX-->

```

```

<din:S_NAD_seller>

```

...

```

</din:S_NAD_seller>

```

```

<din:S_CUX>

```

...

```

</din:S_CUX>

```

```

<!--Група сегментів G_SG2 -->

```

```

<din:G_SG2>

```

...

```

</din:G_SG2>

```

```

<!--Сегменти S_UNH, S_CNT і S_UNT-->

```

...

```

</din:M_ORDERS>

```

Висновки. Запропоновані в ISO/TS 20625: 2002 правила перетворення EDIFACT-повідомлень на XML-документи не єдині. Можна використовувати трохи інші, можливо, коротші XML-схеми, і отже, XML-документи. Тому доречні дослідження на «лаконічність» сукупності потенційних XML-схем, що відповідають дереву граматичного розбору кожного EDIFACT-повідомлення. Цей напрям досліджень видається досить перспективним.

1. ISO/IEC 14662:1997 Information technology - Open-EDI reference model.— <http://www.iso.ch/iso/en/>
2. Webber David. XML/EDI Frameworks - The Executive Summary. The XML/EDI Group, August 1997.- <http://www.bgs-ag.de/en/projects/edifact/start.html>

3. ISO 9735-1:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules (Syntax version number: 4) — Part 1: Syntax rules common to all parts, together with syntax service directories for each of the parts.- <http://www.iso.ch/iso/en/>

4. ISO 9735-2:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules (Syntax version number: 4) — Part 2: Syntax rules specific to batch EDI.— <http://www.iso.ch/iso/en/>
5. ISO 9735-3:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) - Application level syntax rules (Syntax version number: 4) - Part 3: Syntax rules specific to interactive EDI.— <http://www.iso.ch/iso/en/>
6. ISO 9735-4:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules (Syntax version number: 4) — Part 4: Syntax and service report message for batch EDI (message type - CONTRL).- <http://www.iso.ch/iso/en/>
7. ISO 9735-5:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules (Syntax version number: 4) - Part 5: Security rules for batch EDI (authenticity, integrity and non-repudiation of origin).— <http://www.iso.ch/iso/en/>
8. ISO 9735-6:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules (Syntax version number: 4) — Part 6: Secure authentication and acknowledgement message (message type — AUTACK).- <http://www.iso.ch/iso/en/>
9. ISO 9735-7:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules — Part 7: Security rules for batch EDI (confidentiality).- <http://www.iso.ch/iso/en/>
10. ISO 9735-8:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules (Syntax version number: 4) — Part 8: Associated data in EDI.- <http://www.iso.ch/iso/en/>
11. ISO 9735-9:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules (Syntax version number: 4) — Part 9: Security key and certificate management message (message type— KEYMAN).- <http://www.iso.ch/iso/en/>
12. ISO/TS 20625:2002 Electronic data interchange for administration, commerce and transport (EDIFACT) — Rules for generation of XML scheme files (XSD) on the basis of EDI(FACT) implementation guidelines.— <http://www.iso.ch/iso/en/>
13. Календарев А. Построение XML/EDI систем.- <http://www.ciforum.ru>

A. V. Hrechko, D. M. Poshyvanyk

THE GRAMMATICAL MODEL FOR THE CONVERSION OF EDIFACT MESSAGE TO THE XML DOCUMENT

The paper is about building a context-free grammar for checking of EDIFACT messages. The check is based on the syntax notation set specified in ISO 9735. Additionally, a mechanism for syntax tree conversion (of EDIFACT message to the appropriate XML scheme and XML document) is developed on the basis of rules for XML scheme generation according to ISO/TS 20625:2002.