Розділ 1

ВСТУП

Спочатку дамо неформальне означення головного об'єкта нашого курсу – інформації і проблем її захисту.

Безпека діяльності людей потрібна практично для кожного підприємства, організації, банку, торгових мереж тощо. Різниця полягає тільки в тому, які методи будуть потрібні для забезпечення такої безпеки. В загальному випадку згідно з історичною і міжнародною практикою основними об'єктами, які потребують захисту (відповідно до їх пріоритетів), є наступні:

1) люди; 2) інформація; 3) цінності матеріальні.

Якщо пріоритет захисту і безпеки людей є очевидним, то пріоритет захисту інформації над цінностями матеріальними потребує особливого розгляду.

Проблема захисту і безпеки інформації полягає в перетворенні її в такий спосіб, щоб вона не була доступна особам небажаним. Ця проблема є основним завданням науки, яка має назву **криптології**. Історія криптології є ровесником історії природної мови, за допомогою якої виконується передача інформації від одного суб'єкта до іншого. Розвиток писемності призвів до розвитку криптології як науки. На початку ери людини сама писемність була в певному сенсі криптографічною системою, оскільки письмо було доступне лише особам вибраним. І протягом багатьох століть захист інформації використовувався в основному в процесі обміну повідомленнями між державними чиновниками і військовими. В зв'язку з цим коло людей, які займалися захистом повідомлень, було досить обмеженим, а самі методи захисту трималися в секреті. Ситуація змінилася з появою комп'ютерів та комп'ютерних систем, що призвело до нового розвитку криптологічних систем. Тепер обмін інформацією відбувається за допомогою комп'ютерних систем зв'язку, таких як Інтернет. Кожна людина, де б вона не перебувала, може скористатися цими можливостями, а це стало величезною перевагою таких систем. Основними областями діяльності, де використовується комп'ютерний обмін інформацією, є

1. Військова;	2. Дипломатична;	3. Фінансова;
4. Банківська;	5. Комерційна;	6. Промислова;
7. Наукова;	8. Юридична;	9. Медична;
10. Приватна таємниця.		

Головною метою та задачами захисту інформації є:

1. Секретність інформації; 2. Цілісність інформації;
3. Автентичність інформації; 4. Доступність інформації.

До основних напрямків, аспектів, методів та засобів захисту інформації відносять

1. Юридичні, правові;
2. Методично-нормативні;
3. Організаційні;
4. Фізичні;
5. Технічні – захист від витоку по технічних каналах: електромагнітному, оптичному, акустичному, віброакустичному;
6. Стеганографічні;
7. Криптографічні. Методи математичного захисту інформації;
8. Методи квантової криптографії;
9. Морально-етичні норми.

З точки зору реалізації методи 5 – 8 можна поділити на:

– апаратні засоби,
– програмні засоби,
– програмно-апаратні засоби, кожен з яких має свої переваги та недоліки.

1.1. Криптографічний захист інформації

Криптографічний захист інформації – це різновид захисту, який реалізується за допомогою криптографічних перетворень, спеціальних ключових даних з метою приховування та відновлення змісту інформації, підтвердження достовірності, авторства, запобігання несанкціонованому використанню тощо.

Криптографічне перетворення – це перетворення інформації відповідно до певних правил (логічних, математичних) з метою забезпечення функціонування криптографічних протоколів.

Криптографічний ключ – це параметр, який використовується в криптографічному алгоритмі для вибору конкретного криптографічного перетворення; ключі можуть бути таємними або відкритими.

Криптографічний протокол – це послідовність узгоджених дій згідно з деякими правилами, у відповідності до яких відбувається обмін інформацією між сторонами або учасниками протоколу та її перетворення з використанням криптографічних методів та засобів. Простий приклад криптографічного протоколу – це зашифрування та розшифрування повідомлення.

Криптографія – науково-технічна дисципліна, яка вивчає принципи, методи та засоби інформаційних технологій криптографічного захисту інформації, предметом якої є розробка криптографічних систем.

Криптоаналіз – це науково-технічна дисципліна, яка вивчає методи, способи та засоби аналізу криптографічного захисту інформації: криптографічних систем, криптографічних алгоритмів, протоколів з метою їх розкриття без знання таємних клю-

чів, розкриття будови криптосистем, пошук способів несанкціонованого доступу, підробки даних тощо. Криптоаналіз оцінює складність таких способів розкриття (злому) і стійкість криптографічного захисту інформації. Фахівця, який займається криптоаналізом називають **криптоаналітиком**.

Криптологія, за найбільш поширеною сучасною термінологією, об'єднує в собі дисципліни *криптографію* і *криптоаналіз*.

1.1.1 Етапи розвитку криптографічних засобів

1. Криптографія заміни і перестановки (з давнини до середини і кінця XIX століття).

а) Перший шифр перестановки, застосування якого зафіксоване у військовій справі, (Спарта, V ст. до н.е.) – шифр Скитала. В цьому шифрі використовувався круглий барабан (скитал), на який намотувалась стрічка. На намотану стрічку наносився текст повідомлення, а після розмотування стрічки з барабана на ній була зовні випадкова послідовність літер – шифроване повідомлення. Таємним ключем був діаметр барабана. Криптоаналіз шифру Скитала запропонував Арітотель за допомогою барабана змінного діаметру: якщо при намотуванні на нього стрічки з шифрованим повідомленням у деякому місці вгадувались якісь сенсовні частини слів, то цьому місцю відповідав діаметр справжнього барабана.

б) Прикладом шифру заміни є шифр Цезаря – заміна кожної літери повідомлення на літеру циклічно віддалену в алфавіті на фіксоване число позицій (див. підр. "Прості алгоритми підстановки").

2. Застосування телеграфу для шифрування та кодування (з середини XIX ст.).

3. Використання механічних пристроїв (кінець XIX ст. – 20-і роки XX ст.).

4. Електромеханічні пристрої (з 20-х років XX ст. – середина XX ст.). Приклад – електромеханічний ротор (ENIGMA і PURPLE).

5. Електронні машини (з кінця 40-х років XX ст.).

6. Напівпровідникові криптосистеми.

7. Криптосистеми, засновані на мікросхемах.

8. Застосування комп'ютерної техніки для криптографічного захисту.

9. Квантова криптографія.

До епохи Ренесансу криптографія була заняттям непрофесіональним. В ті часи шифри являли собою здебільшого головоломки. Хоча в ту пору були побудовані шифри, які не були зламані навіть через 100 і більше років! За сучаними поняттями, враховуючи залучення до криптоаналізу комп'ютерів, це були слабкі, нестійкі шифри. Велике просування в криптографії сталося після залучення до цієї справи відомих математиків (відродження – Ф. Вієт, Д. Кардано, середні віки Ф. Гаусс, Гольдбах і ін.). Згодом стали створюватись спеціалізовані державні служби шифрування і дешифрувння. Такі служби були створені, наприклад, Кромвелем у Англії, кардиналом Рішельє у Франції, Б. Хмельницьким в Україні, Петром I в Росії.

У 1883 році вийшла книга Керкгоффа "Військова криптографія", в якій вперше були сформульовані певні вимоги до криптосистем, правила побудови, експлуатації, стійкості криптографічних пристроїв. Частина цих правил вважаються актуальними

і в наші дні.

Наукою, у повному розумінні цього слова, криптографія стала визнаватися з 1949 року після появи у відкритому друці статті Клода Шеннона "Теорія зв'язку в секретних системах" [13]. А з 1976 року з появою алгоритмів Діффі і Хеллмана "Нові напрямки в криптографії" [22] почався новий етап у розвитку криптографії – застосування криптосистем з відкритим ключем, яке створило умови успішного розв'язання цілого ряду назрілих проблем криптографічного захисту інформації.

Класифікація сучасних криптосистем. Сучасні криптосистеми поділяються на декілька класів.

1. **Симетричні (одноключові, з секретним ключем)**. Поділяються на *блокові* та *потокові*. У відправника та одержувача повідомлення один і той самий секретний ключ, вони знаходяться у рівних (симетричних) умовах, можуть як зашифрувати повідомлення, так і розшифрувати його за допомогою таємного ключа.

2. **Асиметричні (двохключові, з відкритим ключем, із загальнодоступними ключами).** Стали відомі з 1976 року, активно використовуються на практиці з 1978 року. У найпростішому випадку мають два ключі: один (відкритий) – у відправника для шифрування, і одержувача (секретний) для розшифрування.

3. **Квантові** (знаходяться на етапі експериментів). Квантові криптосистеми ґрунтуються на принципах роботи квантових комп'ютерів.

1.1.2 Квантовий комп'ютер

Квантовий комп'ютер складається із квантово-механічної системи, обов'язково ізольованої від оточуючого середовища таким чином, щоб її поведінкою можна було ззовні керувати так, щоб жодна подія, не пов'язана з процедурами контролю, не в стані змінити цю поведінку.

Модель для такої системи включає:

– простір станів системи, асоційований з ізольованою квантово-механічною системою, є векторним простором над полем комплексних чисел з визначеним скалярним добутком (гільбертів простір);

– стани системи в довільний момент часу, який повністю описується вектором стану, є одиничним вектором в просторі станів;

– зміна стану замкнутої квантово-механічної системи описується тільки унітарним перетворенням, причому стани системи в різні моменти часу зв'язані унітарним перетворенням, яке залежить тільки від моментів часу;

– вимірювання квантової системи складається з набору лінійних операторів, діючих на просторі станів системи, і фактично є проекцією на ортогональні підпростори.

Принцип квантових обчислень ґрунтується на таких кроках.

Виконується послідовність унітарних операцій простого типу над одним, двома або трьома **кубітами**, яка контролюється класичним комп'ютером. В кінці обчислень стан квантового комп'ютера вимірюється, що і дає шуканий результат обчислень.

Ідея квантових обчислень, вперше висловлена Ю.І. Маніним і Р. Фейнманом, полягає в тому, що квантова система із L дворівневих квантових елементів (квантових

бітів, кубітів) має 2^L лінійно незалежних станів і, у відповідності з принципом квантової суперпозиції, простір станів такого квантового регістра є 2^L-мірним гільбертовим простором. Операція в квантових обчисленнях відповідає повороту вектора стану регістра в цьому просторі. Отже, квантовий пристрій обчислень розміру L кубіт фактично задіює одночасно 2^L класичних станів.

За допомогою базових квантових операцій можна симулювати роботу звичних логічних елементів, з яких побудовані класичні комп'ютери. Тому довільну задачу, яка розв'язується на класичних комп'ютерах, квантовий комп'ютер її також розв'яже. Отже, нова схема обчислень буде не слабкішою нинішньої.

Чим же квантовий комп'ютер кращий класичного? Більшість сучасних комп'ютерів працюють за такою схемою: n бітів пам'яті зберігають стан і в кожний такт часу ці стани змінюються процесором. У випадку квантового комп'ютера система з n кубітів знаходиться в стані, який є суперпозицією всіх базових станів і тому зміна в системі стосується всіх 2^n базових станів одночасно. Теоретично нова схема може працювати набагато швидше (в експоненціальне число разів) класичної схеми. Практичний алгоритм пошуку в базі даних показує квадратичний приріст обчислювальної потужності в порівнянні з класичними алгоритмами.

Проблеми квантових обчислень, які виникають на шляху використання квантових комп'ютерів, зводяться до таких:

Декогеренція (нестабільність) при обчисленнях;

Маштабування (мала кількість кубітів);

Компактність;

Дороговизна і доступність.

Практичні досягнення на шляху розвитку квантових комп'ютерів і обчислень:

2018 – 72 qubits, Google, US (Bristlecone 5 березня 2018);

2017 – 50 qubits, IBM, US;

2017 – 17 qubits, IBM, US;

2006 – 12 qubits, Institute for Quantum Computing, Perimeter Institute for Theoretical Physics, and MIT;

2000 – 7 qubits, Los Alamos National Laboratory, US;

2000 – 5 qubits, Technical University of Munich, Germany;

1998 – 2 qubits, IBM, UC Berkeley, Stanford University, and MIT, US;

1998 – 2 qubits, Oxford University, UK.

1.1.3 Основні поняття криптології

Відкритий текст (ВТ) – це повідомлення, дані, елемент простору повідомлень, до якого застосовується процедура криптографічного перетворення, шифрування. Під ВТ розуміють текст, заданий у вигляді послідовності символів скінченого алфавіту, з доступним семантичним змістом. ВТ отримують після правильного розшифрування.

Шифрований текст, шифротекст або криптограма (ШТ) – це інформація, яка одержана після застосування до відкритого тексту процедури зашифрування.

Зашифрування – це криптографічне перетворення відкритого тексту з застосуванням таємних ключів, в результаті якого дістають ШТ або криптограму з недоступним для незаконного користувача семантичним змістом.

Розшифрування – зворотне криптографіне перетворення ШТ з застосуванням таємних ключів, в результаті якого законний користувач дістає ВТ, що був зашифрований. Для даного перетворення також використовується термін *дешифрування*.

Шифратор – пристрій, що здійснює процедуру зашифрування.

Дешифратор – пристрій, що виконує процедуру розшифрування.

Криптографічна система – система безпеки інформації за допомогою криптографічних методів, яка повинна забезпечувати збереження конфіденційності, цілісності, автентичності. З практичної точки зору – це набір апаратних і (або) програмних засобів, інструкцій і правил, за допомогою яких, використовуючи криптографічні перетворення, можна зашифрувати повідомлення і розшифрувати шифрограму різними способами, один з яких вибирається за допомогою секретного ключа, а також виконувати інші криптографічні протоколи. Математичну модель криптографічної системи буде розглянуто далі.

Криптографічна стійкість у широкому розумінні – це здатність криптосистеми або криптоалгоритму протистояти атакам з використанням методів криптоаналізу; у вузькому розумінні (практична стійкість) – це чисельна характеристика часової та ємнісної скадності розкриття криптографічного алгоритму з урахуванням тих науково-технічних методів та засобів, які може використати криптоаналітик.

Теоретична стійкість у широкому розумінні – це стійкість криптосистеми за наявності у криптоаналітика необмеженого часу, необмежених обчислювальних ресурсів, якнайкращих методів криптоаналізу, у вузькому розумінні – це в певному сенсі гарантована стійкість. Основні підходи до визначення поняття теоретичної стійкості нині розглядаються у рамках деяких математичних моделей. Так, розглядається стійкість теоретико-інформаційна, теоретико-складнісна, математично доведена.

Практична стійкість – це стійкість криптосистеми на даний час з урахуванням того, що криптоаналітик володіє обмеженим часом, обмеженими обчислювальними ресурсами і сучасними методами криптоаналізу, а також чисельна характеристика часової та ємнісної складності розкриття криптографічного алгоритму.

1.1.4 Основні види криптографічних атак

Види криптографіяних атак (нападів) залежать від типу інформації, відомої криптоаналітику.

В криптології загально прийняте наступне правило Керкгоффа:

стійкість криптосистеми не повинна опиратися на секретність її будови, алгоритму шифрування, а повинна грунтуватися на секретності ключа (при надійному алгоритмі шифрування і достатньому розмірі ключа).

Залежно від додаткової інформації атаки класифікуються у порядку їх посилення наступним чином.

1. *Атака з використовуванням тільки шифротексту.* У криптоаналітика є ши-

фротексти декількох повідомлень, зашифрованих одним і тим самим алгоритмом шифрування і невідомим ключем (ключами). Задача криптоаналітика полягає в розкритті як можна більшого числа повідомлень або, що краще, отриманні ключа (ключів), використаних для шифрування повідомлень з метою дешифрування також і інших повідомлень, зашифрованих тими ж ключами.

2. *Атака з використанням відкритого тексту.* У криптоаналітика є доступ не лише до шифротекстів декількох повідомлень, але і до відповідних відкритих текстів цих повідомлень. Його завдання полягає в отриманні ключа (або ключів), використаного (використаних) для шифрування повідомлень з метою дешифрування інших повідомлень, зашифрованих тим же ключем (ключами).

3. *Атака на основі вибраного відкритого тексту.* У криптоаналітика є не тільки доступ до шифротекстів і відповідних відкритих текстів декількох повідомлень, але й можливість вибирати відкритий текст (тексти) і отримати шифрований (шифровані). Це надає більше варіантів, ніж атака з використанням відкритого тексту, оскільки криптоаналітик буде вибирати відкриті тексти зі спеціальними властивостями, що може надати більше інформації про ключ. Його завдання полягає в отриманні ключа (або ключів), використаного для шифрування повідомлень, або алгоритму, що дозволяє дешифрувати нові повідомлення, зашифровані тим же ключем (або ключами).

4. *Адаптивна атака з використанням відкритого тексту.* Криптоаналітик не тільки може вибирати тексти для шифрування, але також може будувати свій подальший вибір текстів на базі одержаних результатів шифрування попередніх. При розкритті з використанням вибраного відкритого тексту криптоаналітик може вибрати для шифрування тільки один або декілька ВТ одночасно для отримання відповідних ШТ, при адаптивному нападі з використанням вибраного відкритого тексту він може вибрати спочатку один ВТ і отримати криптограму, потім вибрати наступний ВТ, використовуючи результати першого вибору та шифрування, і так далі. Атаки 2-4 можливі, наприклад, при шифруванні з відкритим ключем.

5. *Атака на основі вибраного шифротекста.* Криптоаналітик має можливість вибирати різні шифротексти для розшифрування і має доступ до розшифрованих відкритих текстів (наприклад, криптоаналітик має доступ до шифрувального апарату).

6. *Адаптивна атака на основі вибраного шифротекста.* Криптоаналітик має можливість для ШТ, що послідовно вибираються з урахуванням попередніх результатів, отримувати відповідні ВТ (аналогічно п.4). Задача знайти таємний ключ, або дешифрувати повідомлення. Атаки п. 5, 6 особливо небезпечні для криптосистем з відкритим ключем.

7. *Атака на основі вибраного тексту* – об'єднує можливості атак п.3, п.5.

8. *Адаптивна атака на основі вибраного тексту* – об'єднує можливості атак п.4, п.6.

Атаки в цьому списку з більшим номером загалом сильніші і небезпечніші ніж з меншим. Для всіх сучасних шифраторів обов'язкова вимога – стійкість до атаках типу 1 і 2. Якщо у криптоаналітика є деяка додаткова інформація про ключі (окрім їх довжини) або про зв'язок між різними ключами, то напади на криптосистему стають ще небезпечнішими. До такої атаки належить, зокрема, атака з використанням інфор-

мації з так званого побічного каналу, що міститься, наприклад, в електромагнітному випромінюванні шифратора.

Для ілюстрації введених понять та понять, які використовуються в задачах критографії, розглянемо приклад нескладної криптографічної задачі.

Задача. Аліса і Боб хочуть провести вечір разом, але не можуть вирішити куди йти: в кінотеатр чи в театр. З метою знаходження відповіді на це питання вони вирішили провести жеребкування шляхом *підкидання монети* (цей спосіб добре відомий всім).

Аліса бере монету і говорить Бобу: "Ти вибираєш сторону монети, а я її підкидаю". Боб згоден і Аліса підкидає монету. Потім вони обоє дивляться на результат підкидання. Якщо результат той, що вибрав Боб, то він вирішує куди йти, інакше це вирішує Аліса.

Вищеописана процедура обміну інформацією між Алісою і Бобом являє собою багатосторонню гру, яка має спеціальну назву – **протокол**. Протокол – це точно визначена послідовність дій, які виконуються декількома учасниками. Наявність декількох учасників є суттєвою умовою, оскільки коли вся процедура виконується одним учасником, то її назвати протоколом не можна.

Задача (продовження). Уявімо тепер, що Аліса і Боб виконують наведений в задачі протокол по телефону. Аліса пропонує Бобу: "Вибери сторону монети, а я її підкину і скажу хто виграв." Очевидно, що Боб не погодиться на такі умови, тому що він не в змозі перевірити результат жеребкування.

Як мусить поступити Боб в цій ситуації?

Розв'язати проблему Бобу допомагає криптографія. Будемо розглядати результат жеребкування як значення функції $f : \mathcal{Z} \to \mathcal{Z}$, яка визначена на множині цілих чисел $\mathcal{Z}$ зі значеннями в цій же множині і яка має такі властивості.

Властивість 1. *Для функції f*

а) *і кожного числа $x \in \mathcal{Z}$ можна легко обчислити значення $f(x)$, але за значенням $f(x)$ неможливо знайти її аргумент x;*

б) *неможливо знайти пару чисел (x, y) таких, що $x \neq y$ і $f(x) = f(y)$.*

В умові а) використовуються слова "легко" і "неможливо". Сенс цих слів буде строго означений далі засобами теорії складності обчислень [14], а тут ці слова виражають рівень складності певної дії. Припустимо, що Аліса і Боб погодилися використовувати функцію $f(x)$ і що парне число означає ОРЕЛ, а непарне – РЕШКА. Тепер вони готові виконати такий протокол – протокол 1.

Протокол 1 (підкидання монети по телефону)

Початкові умови:

а1) функція $f(x)$ має властивість 1.

а2) якщо x – парне, то $f(x)$ означає ОРЕЛ, інакше – РЕШКА.

1. Аліса вибирає велике випадково згенероване ціле число x, обчислює значення $f(x)$ і повідомляє це значення Бобу по телефону.

2. Боб повідомляє свою догадку про число x: парне чи непарне.

3. Аліса повідомляє Бобу число x.
4. Боб перевіряє значення $f(x)$ і переконується у правильності своєї догадки.

Очевидно, що протокол 1 цілком працездатний. Але при цьому виникає низка питань, одне з яких є таким: "Чи надійний цей протокол?" Це питання в криптографії пов'язане з питанням "аналізу стійкості протоколу". Друге питання "Чи достатньо потужний генератор випадкових чисел, який використовує Аліса?" Якість такого генератора надзвичайно важлива для багатьох серйозних застосувань, де необхідно приймати правильні рішення.

Простий аналіз стійкості протоколу 1 показує наступне.

По-перше, властивість б) функції $f(x)$ означає, що Аліса не може знайти два різні числа x і y, одне з яких парне, а друге непарне. Отже, повідомивши Бобу значення $f(x)$ (крок 1), Аліса фіксує свій вибір числа x і не може від нього відмовитися. На цьому жеребкування закінчується.

По-друге, завдяки властивості а) функції $f(x)$, Боб не може визначити число x, яке відоме Алісі. Отже, припущення Боба на кроці 2 протоколу 1 дійсно є догадкою, оскільки в нього відсутня інформація про число x. На цьому кроці Аліса може переконати Боба в правильності його догадки, назвавши число x (крок 3). Боб може перевірити це самостійно, знайшовши значення $f(x)$ (крок 4) і порівнявши його з числом, яке повідомила Аліса.

Зауважимо, що жеребкування вважається коректним, якщо число x вибирається з достатньо великої множини, так щоб ймовірність вгадати парність числа x не перевищувала 50 процентів.

Аналіз протоколу 1 показує, що багато чого в ньому спрощено і багато деталей не враховано. В результаті наведена версія протоколу 1 не є досконалою. Деякі деталі аналізу протоколів будуть розглянуті далі, тому що для цього необхідно познайомитися з певними математичними поняттями і методами.

Таким чином, структура ~~нашого~~ курсу схематично приймає такий вигляд:

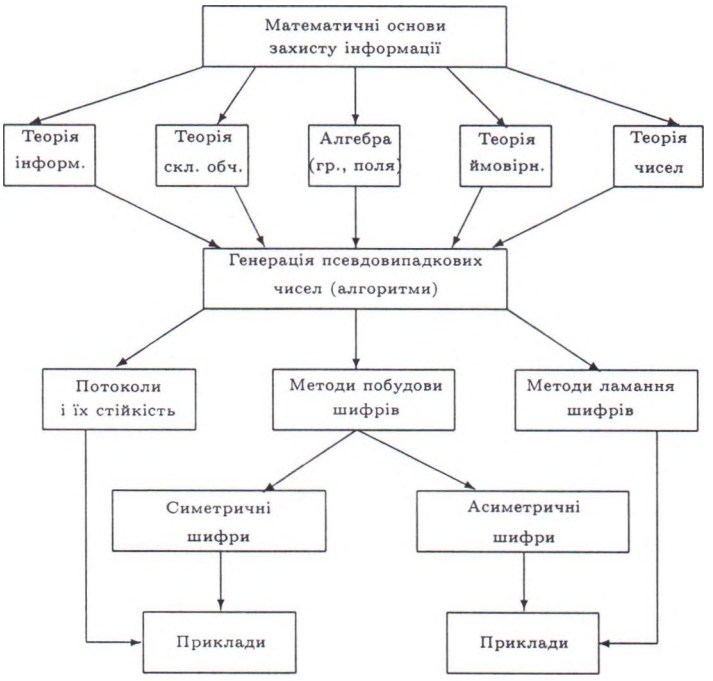


Рис. 1. Схема курсу

За цією схемою і буде йти вивчення математичних основ захисту інформації.

Контрольні питання

1. Які мета та задачі захисту інформації?
2. Назвіть напрямки та методи захисту інформації.
3. Що таке криптографічний ключ?
4. З яких частин складається криптологія?
5. Чи були відомі способи захисту інформації до нашої ери?
6. Яка дата вважається початком розвитку криптографії з відкритим ключем?
7. Яка різниця між симетричними та асиметричними криптосистемами?

Розділ 2

МАТЕМАТИЧНІ ОСНОВИ

2.1. Множини

Інтуїтивне поняття множини. Як відомо, поняття множини відноситься до аксіоматичних понять математики і точне його означення дати неможливо. Часто приймається означення інтуїтивного поняття множини Георга Кантора – основоположника цієї теорії.

Довільне зібрання певних предметів нашої інтуїції чи інтелекту, які можна відрізнити один від одного і які уявляються як єдине ціле, називається множиною. Предмети, які входять до складу множини, називаються її елементами.

З цього означення випливає, що довільна множина повністю визначається своїми елементами. Отже, дві множини рівні, якщо кожний елемент однієї з них є елементом другої і навпаки.

Множина називається **скінченною**, якщо вона складається зі скінченного числа елементів. Запис $a \in A$ ($a \notin A$) означає, що a є (не є) елементом множини A. Та однозначно визначена множина, елементами якої є $a_1, a_2, \ldots, a_n$, позначається $\{a_1, a_2, \ldots, a_n\}$.

Подання множини за допомогою фігурних дужок з явним переліком її елементів можливе лише тоді, коли множина має невелику кількість елементів. Якщо ж множина має хоча й скінченну, але велику кількість елементів, то таке подання множини стає досить громіздким, а у випадку нескінченної множини його застосування взагалі стає неможливе. Виникає питання: як задавати множини, які мають велике або нескінченне число елементів?

Відповідь пов'язана з поняттям властивості або висловлювання. Поки що обмежимося інтуїтивним поняттям висловлювання.

Під **висловлюванням** відносно предмета x будемо розуміти таке розповідне речення, в якому щось стверджується відносно предмета x і яке можна характеризувати як істинне або хибне по відношенню до x.

Приклад 2.1.1. Властивостями є такі записи:

1) 3 ділить 10; 2) $x < x$; 3) $x^2 = 2$; 4) $x^2 + 1 > 0$.

Наведені нижче вирази не є властивостями:

5) стійте тут; 6) існує таке x, що $2x < y$,

тому що їх не можна характеризувати як істинні чи хибні. ♠

Нехай $P(x)$ означає деяку властивість, тоді $P(a)$ буде означати ту ж саму властивість, але з заміною x на a. Задання множини в термінах властивостей досягається за допомогою такого принципу згортання.

Довільна властивість $P(x)$ визначає деяку множину A за допомогою умови: елементами множини A є ті і тільки ті предмети a, які мають властивість P.

Отже, довільна властивість $P(x)$ визначає єдину множину, яку позначають $\{a|\ P(a)\}$ і читають так: "множина всіх тих предметів a, що $P(a)$". Зауважимо, що властивість P може бути способом побудови елементів множини $\{a|\ P(a)\}$.

Безпосередньо з принципу згортання випливає існування такої множини. Нехай A – деяка множина, а $P(x)$ має вигляд $x \neq x$, тоді множина $\{a \in A|P(a)\} = \{a \in A|a \neq a\}$, очевидно, не має елементів. Із принципу об'ємності випливає, що може існувати лише одна множина, яка не має елементів. Ця множина називається **пустою множиною** і позначається $\emptyset$.

2.1.1. Операції, підмножини, алфавіти, мови

Введемо символи $\Leftrightarrow, \exists x, \forall x, \Rightarrow$, які надалі будуть служити для скорочення виразів *"тоді і тільки тоді, коли", "існує x такий, що", "для кожного x" і "випливає"* відповідно.

Основними операціями на множинах є об'єднання, перетин, різниця та її окремий випадок – доповнення.

Означення 1. Об'єднанням $A \cup B$ множин A і B *називається множина, до складу якої входять ті і тільки ті елементи, які входять до складу хоча б однієї із цих множин, тобто $A \cup B = \{a|a \in A$ або $a \in B\}$.*

Перетином $A \cap B$ множин A *і* B *називається множина, елементами якої є елементи, що входять до складу як множини A, так і множини B, тобто $A \cap B = \{a|a \in A$ і $a \in B\}$. Якщо $A \cap B = \emptyset$, то множини A і B називаються такими, що* **не перетинаються**.

Різницею $B \setminus A$ множин B *і* A *називається множина $\{a|a \in B$ і $a \notin A\}$. Очевидно, що $B \setminus A = B \setminus (A \cap B)$.*

Якщо $A \subseteq B$, то $B \setminus A$ називається **доповненням** *множини A в множині B і позначається A'_B або просто A', коли B можна визначити із контексту.*

Симетричною різницею $A \div B$ множин A *і* B *називається множина $(A \setminus B) \cup (B \setminus A)$.*

З означення операції $A \div B$ очевидним чином випливає така рівність: $A \div B = (A \cup B) \setminus (A \cap B)$.

Означення 2. *Множина A називається* **підмножиною** *множини B ($A \subseteq B$), якщо всі елементи множини A є також елементами множини B ($A \subseteq B \Leftrightarrow (a \in A \Rightarrow a \in B)$). При цьому множина B називається* **надмножиною** *множини A.*

Тепер принцип об'ємності можна записати так:

$$A = B \Leftrightarrow (A \subseteq B \text{ і } B \subseteq A).$$

Означення 3. *Введемо знак строгого включення $\subset$ для множин. $A \subset B$ означає для множин A і B, що $A \subseteq B$ і A не рівна B ($A \neq B$). Якщо $A \subset B$, то множина A називається* **власною підмножиною** *множини B, а множина B -* **власною надмножиною** *множини A.*

Введемо до розгляду деякі спеціальні множини.

1) **Алфавітом** X називається скінченна непуста множина попарно різних між собою елементів, які називаються **літерами**: $X = \{x_1, x_2, x_3, \ldots, x_n\}$. **Словом** в алфавіті X називається довільна скінченна послідовність літер цього алфавіту. В криптографії слова довжини n називаються n-грамами, наприклад, якщо $n = 2$ – біграмами, якщо $n = 3$ – триграмами і т. д.

Множина всіх слів скінченної довжини у алфавіті X позначається $F(X)$.

Якщо $p \in F(X)$ – слово, то його **довжиною** називається кількість літер алфавіту X, з яких воно складається, рахуючи кожне входження літери. Будемо позначати довжину слова p через $l(p)$. Наприклад, якщо $X = \{a, b\}$, то $l(aab) = 3$, $l(baabb) = 5$.

Якщо $p = y_1 y_2 \ldots y_k$ і $q = z_1 z_2 \ldots z_r$ – слова в алфавіті X, то $p = q$ тоді і тільки тоді, коли $k = r$ і $y_i = z_i$ для всіх $i = 1, 2, \ldots k$.

Довільна підмножина множини $F(X)$ називається **мовою в алфавіті** X. Наприклад,

1. нехай $X = \{a, b, c, d, ..., z\}$ складається з 26 літер англійського алфавіту. Довільна послідовність літер з X належить до $F(X)$. Отже, $F(X)$ включає слова *math*, *is*, *fun*, *men*, *amour* і т. д. Оскільки $F(X)$ включає множину слів $a, aa, aaa, aaaa, aaaaa, ...$, то $F(X)$ є нескінченною множиною. Можна було б визначити американську мову L як підмножину $F(X)$, що складається зі слів сучасного видання *Webster's New Word Dictionary of the American Language*. Приклад цієї мови показує, що хоча L включає велику кількість елементів, але вона є скінченною множиною.

2. якщо $X = \{0, 1\}$, то множина B тих слів з множини $F(X)$, які починаються з 1 являє собою множину двійкових (бінарних) цілих додатніх чисел, тобто $B = \{1, 10, 11, 100, 101, 110, 111, 1000, 1001, ...\}$.

Введемо до множини $F(X)$ деяке спеціальне слово, яке певною мірою аналогічне до пустої множини і яке називають *пустим словом*: пусте слово за означенням не включає жодного символу алфавіту X і позначається через e (отже, довжина $l(e) = 0$). Наприклад,

a) якщо $X = \{a\}$, то $F(X) = \{e, a, aa, aaa, aaaa, aaaaa, ...\}$;

b) якщо $X = \{a, b\}$, то $F(X) = \{e, a, b, aa, ab, ba, bb, aaa, aab, ...\}$;

c) якщо $X = \{0, 1, 2\}$, то $F(X) = \{e, 0, 1, 2, 00, 01, 02, 10, 11, 12, 20, 21, 22, 000, ...\}$.

Конкатенацією двох слів p і q в алфавіті X називається слово pq, одержане в результаті приписування праворуч до слова p слова q. Очевидно, що для пустого слова справедливі такі рівності: $\forall p \in F(X)(ep = pe = p)$.

Нехай дано слово $q = uw$, де $u, w \in F(X)$. Слово u називається **початком** або **префіксом** слова q, а слово w – **кінцем** або **суфіксом** слова q. Слово p є **підсловом** слова q в алфавіті X, якщо

$$q = p'pq',$$

де p', q' деякі підходящі слова з $F(X)$. Якщо p' є словом найменшої можливої довжини, то входження слова p в слово q називається **першим**. Аналогічно можна означити друге входження слова, третє і т. д. Наприклад, якщо $X = \{a, b\}$, то

$$L = \{p \in F(X) | l(p) = 2\} = \{aa, ab, ba, bb\};$$

якщо $L' = \{p \in F(X) | l(p)$ є числом парним $\}$, то

$$L' = \{aa, ab, ba, bb, aaaa, aaab, aabb, abbb, ...\}.$$

Зазначимо, що L є підмножиною множини L'. Слово $p = baac$ є підсловом слова $q = abaacb$. Слова $e, a, ab, aba, abaa, abaac, abaacb$ є префіксами слова q, а слова $e, b, cb, acb, aacb, baacb, abaacb$ є суфіксами слова q.

2) Множина натуральних чисел $\{0, 1, 2, \ldots, n, \ldots\}$, яку позначають буквою $\mathcal{N}$, задається за допомогою принципу згортання з використанням поняття мови. Нехай $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ – множина з десяти елементів, яка складає алфавіт. Словами в цьому алфавіті є довільна скінченна послідовність символів із A. Наприклад, $p = 0020034985700$ – слово в алфавіті A, а $p = 00a26543c$ не є словом в алфавіті A, оскільки до його запису входять символи (символи a, c), що не належать множині A. Тобто p буде словом в алфавіті A тоді і тільки тоді, коли кожний символ цього слова належить алфавіту A.

Тоді множина $\mathcal{N} = \{p = xy...z | x, y, ..., z \in A\}$ складає множину натуральних чисел. Зауважимо, що слово $p = 00...0$ і $p' = 0$ – одне і те саме число 0, так само, як і числа 00057 і 57. Отже, запис одного і того самого елемента множини $\mathcal{N}$ може бути різним. Для однозначності запису чисел необхідно поставити вимогу, щоб перший символ у слові p, яке складається більше ніж з одного символа, був відмінний від нуля, тобто $\mathcal{N} = \{p = xy \ldots z | x, y, \ldots, z \in A$ і $x \neq 0$ або 0, якщо $p = 0\}$.

3) Множина натуральних додатних чисел $\mathcal{N}^+ = \{1, 2, \ldots, n, \ldots\}$ – це така множина слів в алфавіті A, серед яких немає слова вигляду $p = 0$.

4) Множина цілих чисел $\mathcal{Z} = \{\ldots, -n, \ldots, -2, -1, 0, 1, 2, \ldots\ldots, n, \ldots\}$ – це множина слів у алфавіті $B = \{-, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ таких, що $\mathcal{Z} = \{p = xy...z | x, y, ..., z \in B$ і $x \neq 0$ і $y, ..., z \neq -$ або 0, якщо $p = 0\}$. Слово, першим символом якого є символ "-", називається від'ємним, а слово, першим символом якого є символ, відмінний від символа "-", називається додатним.

5) Множина раціональних чисел $\mathcal{Q}$ складається із множини всіх цілих чисел і всіх нескорочуваних раціональних дробів виду m/n, де $m, n \in \mathcal{Z}, n \neq 0$, тобто $\mathcal{Q} = \{m/n | m, n \in \mathcal{Z} \ i \ n \neq 0\}$.

6) Множина дійсних чисел $\mathcal{D}$ складається із множини всіх раціональних чисел і всіх ірраціональних чисел (чисел, які подаються у вигляді нескінченних неперіодичних десяткових дробів).

7) Для введених вище множин маємо такі включення $\mathcal{N}^+ \subset \mathcal{N}, \mathcal{N} \subset \mathcal{Z}, \mathcal{Z} \subset \mathcal{Q}, \mathcal{Q} \subset \mathcal{D}$ або $\mathcal{N}^+ \subset \mathcal{N} \subset \mathcal{Z} \subset \mathcal{Q} \subset \mathcal{D}$.

Зазначимо, що для довільної множини S із $x \in S$ випливає $x \in S$ і тому $S \subseteq S$. Це означає, що довільна множина є своєю підмножиною. Для того щоб розрізняти підмножини і власні підмножини і вводилися позначення $\subseteq$ і $\subset$.

Покажемо, що пуста множина є підмножиною довільної множини A. Припустимо, що $\emptyset \subseteq A$ хибне. Це можливо лише тоді, коли існує хоча б один елемент із множини $\emptyset$, який не є елементом множини A. Але це неможливо, оскільки пуста множина не має елементів взагалі. Отже, $\emptyset \subseteq A$.

Ще одна важлива множина часто зустічається в застосуваннях – це булеан множини: *для будь-якої множини U існує множина $B(U)$ всіх підмножин множини U*. Множину $B(U)$ також називають **множиною-степенем** множини U.

Зазначимо, що коли множина S має n елементів, то її булеан $B(S)$ має 2^n елементів. Якщо множина S нескінченна, то очевидно, що множина $B(S)$ теж буде нескінченною.

2.2. Відношення і функції

При вивченні властивостей множини не розглядалися зв'язки між елементами однієї і тієї множини або різних множин. Але в довколишній дійсності всі об'єкти, як елементи деяких множин, пов'язані між собою певними відношеннями. Для строгого математичного означення поняття відношення необхідно ввести операцію декартового добутку множин.

Нехай A і B – дві множини. Розглянемо множину упорядкованих пар вигляду

$$C = \{(a, b) | a \in A, b \in B\}.$$

Ця множина називається **декартовим добутком множин** A і B і позначається $A \times B$. Якщо множини A і B скінченні та складаються відповідно із m і n елементів, то очевидно, що C складається із $m \cdot n$ елементів.

Самостійний інтерес являє собою випадок, коли множини A і B рівні між собою: $A = B$. Для його розгляду введемо поняття упорядкованої пари.

Упорядкованою парою елементів множини A будемо називати об'єкт (a, a'), що складається з двох, не обов'язково різних, елементів a і a' множини A, для яких вказано, котрий із них потрібно вважати першим, а котрий – другим. Множина $C = \{(a, a') | a, a' \in A\}$ всіх упорядкованих пар із множини A називається **декартовим квадратом** множини A і позначається A^2.

Наведені означення декартового добутку двох множин і декартового квадрата множини узагальнюється і на випадок довільної скінченної сукупності множин.

Означення 4. Декартовим добутком $A_1 \times A_2 \times \ldots \times A_n$ *множин* $A_1, A_2, \ldots, A_n$ *називається сукупність послідовностей (сукупність упорядкованих* n*-ок елементів) виду* $(a_1, a_2, \ldots, a_n)$*, де* $a_i \in A_i, 1 \leq i \leq n$.

Означення 5. *Довільна підмножина* R *множини* $A_1 \times A_2 \times \ldots \ldots \times A_n$ *називається* **відношенням**, *заданим або визначеним на множинах* $A_1, A_2, \ldots, A_n$. *Якщо* $A_1 = A_2 = \ldots = A_n = A$, *то декартовий добуток* $A_1 \times A_2 \times \ldots \times A_n$ *називається* **декартовим добутком** n**-го степеня множини** $A(A^n)$, *а відношення* R, *задане на множинах* $A_1, A_2, \ldots, A_n$, - n**-арним відношенням** *на множині* A.

Зокрема, при $n = 1$ відношення називається *унарним*, при $n = 2$ відношення називається *бінарним*, при $n = 3$ - *тернарним*, при $n = 4$ - *чотириарним*, при $n = 5$ - *п'ятиарним* і т. д.

Коли $(a_1, a_2, \ldots, a_n) \in R$, то говорять, що елементи a_i $(i = 1, 2, \ldots, n)$ знаходяться у відношенні R між собою або що відношення R **істинне** для $a_1, a_2, \ldots, a_n$. Якщо $(a_1, a_2, \ldots, a_n) \notin R$, то говорять, що R **хибне** для $a_1, a_2, \ldots, a_n$. Зокрема, якщо $R = A_1 \times A_2 \times \ldots \times A_n$, то відношення R називається **тотожно істинним**, а коли $R = \emptyset$, то відношення R називається **тотожно хибним**.

2.2.1. Функції, операції, предикати

Відношення F, задане на множинах $A_1, A_2, \ldots, A_n, B$, називається **функціональним,** якщо для довільного елемента $(a_1, a_2, \ldots, a_n) \in A_1 \times A_2 \times \ldots \times A_n$ існує не більше одного елемента b із B такого, що $(a_1, a_2, \ldots, a_n, b) \in F$. Якщо такий елемент b із B існує для деякого $(a_1, a_2, \ldots, a_n)$, то він позначається через $F(a_1, a_2, \ldots, a_n)$ і записується так: $b = F(a_1, a_2, \ldots, a_n)$.

Нехай $F^{-1}(b) = \{(a_1, \ldots, a_n) \in A_1 \times \ldots \times A_n | F(a_1, \ldots, a_n) = b\}$ і

$$Dom(F) = \bigcup_{b \in B} F^{-1}(b).$$

Очевидно, що для довільного функціонального відношення F, заданого на $A_1, A_2, \ldots, A_n, B$, має місце включення

$$Dom(F) \subseteq A_1 \times A_2 \times \ldots \times A_n.$$

Означення 6. *Відношення* F *називається* **повністю визначеним**, *якщо* $Dom(F) = A_1 \times A_2 \times \ldots \times A_n$, *і* **частково визначеним** *або просто* **частковим**, *якщо* $Dom(F) \subset A_1 \times A_2 \times \ldots \times A_n$.

Відношення F, *задане на множинах* $A_1, A_2, \ldots, A_n, B$, *називається* **відображенням**, *або* **функцією** *із* $A_1 \times A_2 \times \ldots \times A_n$ *в* B $(F : A_1 \times A_2 \times \ldots \times A_n \to B)$, *якщо* F *функціональне і повністю визначене. Відношення* F *називається* **частковим відображенням**, *або* **частковою функцією**, *якщо* F *функціональне і часткове. Число* n *називається* **арністю** *відображення* F.

Якщо $F : A_1 \times A_2 \times \ldots \times A_n \to B$ і існує b із B такий, що $F(a_1, a_2, \ldots, a_n) = b$, то елемент b називають *образом* елемента $(a_1, a_2, \ldots, a_n)$ при відображенні F, а елемент $(a_1, a_2, \ldots, a_n)$ – *прообразом* елемента b. Множину

$$F^{-1}(b) = \{(a_1, a_2, \ldots, a_n) | F(a_1, a_2, \ldots, a_n) = b\},$$

введену раніше, називають *повним прообразом* елемента b в множині $A_1 \times A_2 \times \ldots \times A_n$.

Нехай $F : A_1 \times A_2 \times \ldots \times A_n \to B$ і $F^{-1}(b)$ – повний прообраз елемента b в $A_1 \times A_2 \times \ldots \times A_n$ при відображенні F. Введена вище множина $Dom(F) = \bigcup\limits_{b \in B} F^{-1}(b)$ називається **областю визначення** відображення F, а $Im(F) = \{b | F^{-1}(b) \neq \emptyset\}$ – **областю значень.** Якщо $F : A_1 \times A_2 \times \ldots \times A_n \to B$ і $F_1 : A_1 \times A_2 \times \ldots \times A_n \to B$, то $F = F_1 \Leftrightarrow Dom(F) = Dom(F_1), Im(F) = Im(F_1)$ і $F(a_1, a_2, \ldots, a_n) = = F_1(a_1, a_2, \ldots, a_n)$ для довільної n-ки $(a_1, a_2, \ldots, a_n) \in A_1 \times A_2 \times \ldots \times A_n$.

Означення 7. *Відображення $\varepsilon : A_1 \times A_2 \times \ldots \times A_n \to A_1 \times \times A_2 \times \ldots \times A_n$, яке ставить у відповідність кожному елементові множини $A_1 \times A_2 \times \ldots \times A_n$ той самий елемент, тобто $\forall (a_1, a_2, \ldots, a_n) \in A_1 \times A_2 \times \ldots \times A_n$ $\varepsilon(a_1, a_2, \ldots, a_n) = (a_1, a_2, \ldots, a_n)$, називається* **тотожним відображенням**.

Відображення $F : A_1 \times A_2 \times \ldots \times A_n \to B$ називають **вкладенням**, *або* **ін'єкцією**, *тоді і тільки тоді, коли із $(a_1, a_2, \ldots, a_n) \neq \neq (a'_1, a'_2, \ldots, a'_n)$ випливає $F(a_1, a_2, \ldots, a_n) \neq F(a'_1, a'_2, \ldots, a'_n)$.*

Відображення $F : A_1 \times A_2 \times \ldots \times A_n \to B$ називають **відображенням "на"**, *або* **сюр'єкцією**, *тоді і тільки тоді, коли $(\forall b \in B)(F^{-1}(b) \neq \emptyset)$.*

Відображення F множини $A_1 \times A_2 \times \ldots \times A_n$ на множину B називається **взаємно однозначним відображенням**, *або* **взаємно однозначною відповідністю**, *або* **бієкцією**, *тоді і тільки тоді, коли воно є ін'єкцією і сюр'єкцією одночасно.*

Якщо $f : A \to B$ – відображення і $A_1 \subseteq A$, то відображення $f_{A_1} : A_1 \to B$ називається **звуженням**, *або* **обмеженням**, *відображення f з множини A на множину A_1, коли $f_{A_1}(a) = f(a)$ для всіх $a \in A_1$.*

Якщо $A_1 \subseteq A$ і $f : A_1 \to B$ – відображення, то відображення $g : A \to B$ називається **розширенням** *відображення f з множини A_1 на множину A, коли $g(a) = f(a)$ для всіх $a \in A_1$.*

Якщо $f : A \to B$ – відображення, а множини A і B – скінченні і мають відповідно m та n елементів, то необхідною умовою ін'єктивності f є умова $m \leq n$, сюр'єктивності f – умова $n \leq m$, бієктивності f – умова $m = n$.

Приклад 2.2.1. Відображення відіграють основну роль в процесі передачі та обробки інформації. Це пов'язано, перш за все, з поняттям кодування.

Нехай $X = \{x_1, x_2, \ldots, x_r\}$ і $Y = \{y_1, y_2, \ldots, y_m\}$ – два скінченні алфавіти.

Алфавітним відображенням називається відображення

$$\varphi : X \to F(Y),$$

таке, що $\varphi(x_i) = q_i$, де $i = 1, 2, \ldots r$, а $F(Y)$ – множина всіх слів скінченної довжини в алфавіті Y. При цьому відображення φ розширюється на множину слів $F(X)$ таким чином: $\varphi(x_{i_1} x_{i_2} \ldots x_{i_s}) = \varphi(x_{i_1})\varphi(x_{i_2}) \ldots \varphi(x_{i_s}) = q_{i_1} q_{i_2} \ldots q_{i_s}$. Алфавіт Y в такому випадку називається кодуючим, а множина слів $\{q_1, q_2, \ldots, q_r\}$ – **множиною кодових слів**, або просто **кодом**.

Кодування φ називається **взаємно однозначним**, якщо для довільних $p_1, p_2 \in F(X)$ із $p_1 \neq p_2$ випливає $\varphi(p_1) \neq \varphi(p_2)$.

Код називається префіксним (суфіксним), якщо жодне з кодових слів не є початком (кінцем) іншого слова. Неважко показати, що довільний префіксний код є взаємно однозначним кодом.

Найчастіше в застосуваннях кодуючим алфавітом виступає двійковий алфавіт $Y = \{0, 1\}$. ♠

Вище була введена операція добутку відношень. Оскільки відображення – це відношення спеціального виду, то з'ясуємо, що собою являє добуток відображень.

Нехай $F_1 : A \to B$, а $F : B \to C$ – деякі відображення. З означення операції добутку відношень маємо: $(a, c) \in F_1 * F \Leftrightarrow$ існує елемент $b \in B$ такий, що $(a, b) \in F_1$ і $(b, c) \in F$, тобто $F_1(a) = b$ і $F(b) = c$, або $F(F_1(a)) = c$ згідно з прийнятими вище позначеннями. Таким чином, добуток відображень являє собою добре відому операцію суперпозиції функцій.

Для ін'єкцій, сюр'єкцій та бієкцій має місце

Твердження 1. а) *добуток ін'єктивних відображень є ін'єкцією*;

б) *добуток сюр'єктивних відображень є сюр'єкцією*;

в) *добуток бієктивних відображень є бієкцією.*

Доведення. Не обмежуючи загальності, будемо вважати, що відображення f і f_1 мають вигляд $f : A \to B$ і $f_1 : B \to C$.

а) Нехай $a \neq a'$, тоді на підставі ін'єктивності f отримуємо $f(a) \neq \neq f(a')$. Але тоді на підставі ін'єктивності f_1 дістаємо $f_1(f(a)) \neq \neq f_1(f(a'))$. А це означає ін'єктивність відображення $f * f_1$;

б) На підставі сюр'єктивності відображень f і f_1 для довільного елемента $c \in C$ існує елемент $b \in B$ такий, що $f_1(b) = c$. Далі для довільного елемента $b \in B$ існує елемент $a \in A$ такий, що $f(a) = = b$. Але тоді $f_1(f(a)) = c$, що означає сюр'єктивність відображення $f * f_1$;

в) доведення випливає з попередніх пунктів *а*), *б*). ■

Нехай ε_A і ε_B означають тотожні відображення множин A і B відповідно. Очевидно, що $\varepsilon_A * f = f$ і $f * \varepsilon_B = f$ для довільного відображення $f : A \to B$.

Означення 8. *Відображення $g : B \to A$ називається* **оберненим до відображення** *$f : A \to B$, якщо $f * g = \varepsilon_A$ і $g * f = \varepsilon_B$. Обернене відображення до відображення f позначається f^{-1}. Це означає, що $f(a) = b \Leftrightarrow f^{-1}(b) = a$.*

Приклад 2.2.2. 1) Нехай дано множини $X = \{a, b, c, d\}$, $Y = \{2, 3, 4, 5, 6\}$ і функція f, яка визначена таким чином: $f(a) = 2, f(b) = 4, f(c) = 6, f(d) = 3$. Функцію f можна подати у вигляді такої діаграми, з якої видно що f ін'єкція:

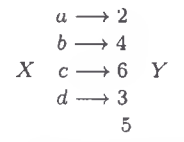


Рис. 2.1.1. Діаграма функції f

Прообразом елемента 4 буде одноелементна множина b, прообразом елемента 5 буде пуста множина $\emptyset$, а множина значень даної функції складається з елементів $\{2, 4, 6, 3\}$.

2) Нехай $X = \{1, 2, 3, \ldots, 10\}$, $Y = X$ і функція f є такою:

$$f(1)=1,\ f(2)=4,\ f(3)=9,\ f(4)=5, f(5)=3,$$
$$f(6)=3,\ f(7)=5,\ f(8)=9,\ f(9)=4, f(10)=1.$$

Для цієї функції маємо таку множину значень $Y=\{1,3,4,5,9\}$. Звідси випливає, що дана функція є повністю визначеною, є відображенням в множину Y, а не на множину Y (елементи 2, 6, 7, 8, 10 мають пусті прообрази в X), а отже, не є взаємно однозначним відображенням.

3) Нехай $X=\{a,b,c,d,e\}$, $Y=\{1,2,3,4,5\}$ і розглянемо функцію f визначену за допомогою такої діаграми:

$$X\ \begin{array}{l}a\longrightarrow 5\\ b\longrightarrow 2\\ c\longrightarrow 1\\ d\longrightarrow 4\\ e\longrightarrow 3\end{array}\ Y \qquad\qquad Y\ \begin{array}{l}5\longrightarrow a\\ 2\longrightarrow b\\ 1\longrightarrow c\\ 4\longrightarrow d\\ 3\longrightarrow e\end{array}\ X.$$

З цієї діаграми випливає, що функція $f: X\to Y$ ін'єкція і, навіть, бієкція, і для цієї функції існує обернена до неї функція $g=f^{-1}: Y\to X$.

4) У випадку нескінченних областей означення і значень функцій встановлення існування обернених функцій вимагає певних зусиль. Розглянемо приклади.

а) Довести, що функція $f:\mathcal{D}\to\mathcal{D}$, задана виразом $f(x)=4x+3$, є бієкцією;

б) Довести, що функція $f:\mathcal{D}\setminus\{1\}\to\mathcal{D}\setminus\{1\}$, задана виразом $f(x)=\frac{x}{x-1}$, є бієкцією. Знайти для неї обернену функцію.

Доведення. а) Припустимо, що $f(a)=f(b)$ для деяких $a,b\in\mathcal{D}$, тобто $4a+3=4b+3$. Звідси отримуємо $4a=4b$ і, отже, $a=b$. Таким чином, функція f є ін'єкцією.

Нехай $b\in\mathcal{D}$ – довільний елемент. Покажемо, що існує $a\in\mathcal{D}$ такий, що $f(a)=b$. Досить прості перетворення дають $a=\frac{1}{4}(b-3)$.

Отже, дана функція є ін'єкцією і сюр'єкцією, і тому є бієкцією.

б) Припустимо, що $f(a)=f(b)$ для деяких $a,b\in\mathcal{D}\setminus\{1\}$, тобто $\frac{a}{a-1}=\frac{b}{b-1}$. Звідси отримуємо $ab-a=ab-b$ і, отже, $a=b$. Таким чином, функція f є ін'єкцією.

Нехай $b\in\mathcal{D}\setminus\{1\}$ – довільний елемент. Покажемо, що існує $a\in\mathcal{D}\setminus\{1\}$ такий, що $f(a)=b$. Досить прості перетворення дають $a=\frac{b}{b-1}$. На підставі довільності елемента b отримуємо, що f – сюр'єкція.

Отже, дана функція є ін'єкцією і сюр'єкцією, а тому є бієкцією.

Обернена функція f^{-1} знаходиться з умови $f^{-1}(b)=a\Leftrightarrow f(a)=b$. Але зі знайденого вище маємо $a=\frac{b}{b-1}$.

Таким чином, $f^{-1}:\mathcal{D}\setminus\{1\}\to\mathcal{D}\setminus\{1\}$ має вигляд $f^{-1}(x)=\frac{x}{x-1}$, тобто f обернена до самої себе.

5) Нехай для $A=\{1,2,3,4,5\}$ відображення $f: A\to\mathcal{D}$, де $\mathcal{D}$ – множина дійсних чисел і $f=\{(1,10),(2,13),(3,16),(4,19),(5,22)\}$, $g:\mathcal{Q}\to\mathcal{D}$, де $g(m)=3m+7$ для всіх $m\in\mathcal{Q}$ і $h:\mathcal{D}\to\mathcal{D}$ – відображення таке, що $h(r)=3r+7$ для всіх $r\in\mathcal{D}$. Тоді

а) g – розширення f з A на $\mathcal{Q}$; б) f – звуження g з $\mathcal{Q}$ на A;

в) h – розширення f з A на $\mathcal{D}$; г) f – звуження h з $\mathcal{D}$ на A;

д) h – розширення g з $\mathcal{Q}$ на $\mathcal{D}$; е) g – звуження h з $\mathcal{D}$ на $\mathcal{Q}$. ♠

Розглянемо питання про умови, за яких існує обернене відображення, та властивості обернених відображень. ~~Має місце~~

Теорема 1. а) *відображення $f: A\to B$ має обернене відображення f^{-1} тоді і тільки тоді, коли f – бієкція;*

б) *якщо відображення f має обернене відображення, то це відображення єдине;*

в) *якщо відображення $f: A\to B$ – бієкція, то обернене відображення f^{-1} теж бієкція і $(f^{-1})^{-1}=f$;*

г) *якщо відображення $f: A\to B$ і $g: B\to C$ – бієкції, то їх добуток $f*g$ теж бієкція і $(f*g)^{-1}=g^{-1}*f^{-1}$;*

д) *якщо множина A скінченна і $f: A\to A$ є ін'єкцією або сюр'єкцією, то f буде бієкцією.*

Доведення. а) Доведемо, що коли $f: A\to B$ і $g: B\to A$ – довільні відображення,

які задовольняють умову $f * g = \varepsilon_A$, то f – ін'єкція, а g – сюр'єкція. Дійсно, якщо $a, a' \in A$ і $f(a) = f(a')$, то

$$a = \varepsilon_A(a) = f * g(a) = g(f(a)) = g(f(a')) = f * g(a') = \varepsilon_A(a') = a'.$$

Отже, відображення f – ін'єкція. Якщо $a \in A$ – довільний елемент, то $a = \varepsilon_A(a) = f * g(a) = g(f(a))$, а це доводить сюр'єктивність відображення g.

Припустимо, що відображення f має обернене f^{-1}. Тоді із $f* *f^{-1} = \varepsilon_A$ і $f^{-1} * f = \varepsilon_B$ випливає, що f сюр'єкція і ін'єкція, тобто f – бієкція.

Навпаки, припустимо, що f – бієкція. Тоді для довільного $b \in B$ знайдеться єдиний елемент $a \in A$, який є прообразом елемента b, тобто $f(a) = b$. Покладаючи $g(b) = a$, визначаємо відображення $g : B \to A$, яке задовольняє умови $f * g = \varepsilon_A$ і $g * f = \varepsilon_B$, Отже, $g = f^{-1}$.

б) Припустимо, що існує два відображення g і g', які обернені до відображення f, тобто $f * g = \varepsilon_A$ і $g * f = \varepsilon_B$ та $f * g' = \varepsilon_A$ і $g' * f = \varepsilon_B$. Тоді отримуємо

$$g' = \varepsilon_B * g' = (g * f) * g' = g * (f * g') = g * \varepsilon_A = g.$$

в) Якщо f бієкція, то за попереднім пунктом a) існує єдине f^{-1}. На тій же підставі відображення f^{-1} теж буде бієкцією. Із симетричності умов $f * f^{-1} = \varepsilon_A$ і $f^{-1} * f = \varepsilon_B$ випливає $(f^{-1})^{-1} = f$.

г) Те, що добуток $f * g$ є бієкцією, випливає з попереднього твердження. А з умови цього пункту і пункту a) теореми випливає існування відображень $f^{-1} : B \to A$ і $g^{-1} : C \to B$, а також їх суперпозиції $g^{-1} * f^{-1} : C \to A$. Звідси отримуємо:

$$(f * g) * (g^{-1} * f^{-1}) = ((f * g) * g^{-1}) * f^{-1} = (f * (g * g^{-1})) * f^{-1} = f * f^{-1} = \varepsilon_A,$$
$$(g^{-1} * f^{-1}) * (f * g) = ((g^{-1} * f^{-1}) * f) * g = (g^{-1} * (f * f^{-1})) * g = g^{-1} * g = \varepsilon_C.$$

А це означає, що відображення $g^{-1} * f^{-1}$ обернене до відображення $f * g$.

д) Нехай f ін'єкція, необхідно показати, що f сюр'єкція. Покладемо для довільного $a \in A$ $f^k(a) = \underbrace{f(f \dots f}_{k \text{ разів}}(a) \dots) = f(f^{k-1}(a))$, $k = 0, 1, 2, \dots$. Оскільки множина A скінченна, то в цій послідовності елементи повинні повторитися. Нехай, наприклад, $f^m(a) = f^n(a)$ і $m > n$. Якщо $n > 0$, то із $f(f^{m-1}(a)) = f(f^{n-1}(a))$ випливає $f^{m-1}(a) = f^{n-1}(a)$. Повторивши це перетворення n разів, скоротимо f і приходимо до рівності $a = f^{m-n}(a) = f(f^{m-n-1}(a))$. А звідси отримуємо елемент $a' = f^{m-n-1}(a)$, для якого $f(a') = f^{m-n}(a) = a$.

Доведення у випадку сюр'єкції f проводиться аналогічно. ■

Розглянемо приклади, які ілюструють введені поняття.

Приклад 2.2.3. Дано функції $f : \mathcal{D} \to \mathcal{D}$ і $g : \mathcal{D} \to \mathcal{D}$, де $f(x) = x^2$, $g(x) = 2x + 1$. Знайти $f * g$, $g * f$, $f * f$ і $g * g$.

Розв'язання. Всі функції визначені на множині дійсних чисел $\mathcal{D}$. Отже,

$(f * g)(x) = g(f(x)) = g(x^2) = 2x^2 + 1$;

$(g * f)(x) = f(g(x)) = f(2x + 1) = (2x + 1)^2 = 4x^2 + 4x + 1$;

$(f * f)(x) = f(f(x)) = f(x^2) = x^4$;

$(g * g)(x) = g(g(x)) = g(2x + 1) = 2(2x + 1) + 1 = 4x + 3$. ♠

Приклади функцій, які часто зустрічаються в застосуваннях.

а) *Селекторні функції.* Функція $I_m^n(x_1, \ldots, x_n) = x_m$ називається **селекторною функцією**, де $1 \leq m \leq n$.

б) *Функція "низ"* – $\lfloor x \rfloor$ – дає найбільше ціле число, яке не перевищує x, де $x \in \mathcal{D}$.

в) *Функція "верх"* – $\lceil x \rceil$ – дає найменше ціле число, яке не менше x, де $x \in \mathcal{D}$.

г) *Функція "остача".* Для довільного цілого числа a і натурального n функція $rest(a, n)$ являє собою остачу від ділення a на n: $rest(a, n) = a - \lfloor \frac{a}{n} \rfloor \cdot n$. Ця функція деколи буде позначатися як $mod(a, n)$.

д) *Функція логарифм.* Існує декілька функцій цього типу, в залежності від основи:

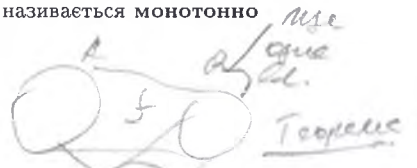
$\lg n = \lg_{10} n$ – десятковий; $\log n = \log_2 n$ – двійковий;

$\ln n = \log_e n$ – натуральний, $\log \log n = \log(\log n)$, $\log^k n = (\log n)^k$.

е) $\chi(a)$ – характеристична функція деякої множини A, значення якої дорівнює одиниці, якщо $a \in A$, і дорівнює нулю, якщо $a \notin A$.

є) *Факторіал* $n!$ натурального числа n покладається для $n = 0$ рівним 1, а для $n > 1$ рівним $n \cdot (n-1)!$.

ж) *Монотонна функція.* Функція $f(n)$ називається **монотонно неспадною (незростаючою)**, якщо із нерівності $m \leq n$ випливає нерівність $f(m) \leq f(n)$ $(f(n) \leq \leq f(m))$. Функція $f(n)$ називається **монотонно спадною (зростаючою)**, якщо із $m < n$ випливає $f(m) > f(n)$ $(f(m) < f(n))$. ♠



2.2.2. Асимптотичне порівняння функцій

При аналізі якісних властивостей алгоритмів важливим є встановлення класу складності алгоритма (строгі означення будуть дані в наступному підрозділі). Складність алгоритму визначається функцією, аргументами якої виступають розміри вхідних даних алгоритма, а її значення на цих даних є його характеристикою. Ця функція називається функцією складності алгоритму, а такий аналіз зводиться до асимптотичного порівняння швидкостей росту функцій.

Існує принаймні чотири категорії алгоритмів за складністю: алгоритми, складність яких росте

– швидше, ніж дана функція $f(n)$;

– не швидше, ніж дана функція $f(n)$;

– з тією ж швидкістю, що і $f(n)$;

– повільніше, ніж перша функція і швидше, ніж друга функція.

Розглянемо категорії росту детальніше. Асимптотична швидкість росту функції являє собою суму складностей окремих частин алгоритма і визначається старшим, домінуючим членом цієї суми. При цьому молодшими членами суми нехтують, оскільки вони ростуть повільніше ніж старший. Нехтуючи молодшими членами, отримуємо те, що називається **порядком росту функції**, який відображає складність алгоритма. В такому випадку аргументи функцій складності, як і їх значення належать множині натуральних чисел. Для ілюстрації розглянемо приклад. Нехай при аналізі алгоритма, встановлено, що він виконує $n^3 + 30n$ порівнювань. Тоді, згідно зі сказаним вище, говоримо, що складність алгоритма росте як n^3. Підставою для такого висновку є те, що вже при розміру вхідних даних $n = 100$ різниця між n^3 і $n^3 + 30n$ складає всього лише 0,3%.

Нехай дано дві функції $f(n)$ і $g(n)$, де значення аргумента n і значення самих функцій належать множині натуральних чисел $\mathcal{N}$.

Означення 9. а) *(О-велике) Вираз* $f(n) = O(g(n))$ *означає, що існують константа* $c > 0$ *і* $n_0 \in \mathcal{N}^+$ *такі, що* $0 \leq f(n) \leq cg(n)$, *для всіх* $n \geq n_0$;

б) (нижня асимптотична границя) *Вираз* $f(n) = \Omega(g(n))$ *означає, що існують константа* $c > 0$ *і* $n_0 \in \mathcal{N}^+$ *такі, що* $0 \leq cg(n) \leq f(n)$, *для всіх* $n \geq n_0$;

в) (пасмова асимптотична границя) *Вираз* $f(n) = \Theta(g(n))$ *означає, що існують константи* $c_1 > 0$ *і* $c_2 > 0$, *де* $c_1 \leq c_2$, *та число* $n_0 \in \mathcal{N}^+$ *такі, що* $0 \leq c_1 g(n) \leq f(n) \leq c_2 g(n)$, *для всіх* $n \geq n_0$;

г) (*о-мале*) *Вираз* $f(n) = o(g(n))$ *означає, що для довільної константи* $c > 0$ *існує число* $n_0 \in \mathcal{N}^+$ *таке, що* $0 \leq f(n) < cg(n)$, *для всіх* $n \geq n_0$.

Якщо $f(n) = \Omega(g(n))$, то це означає, що функція $f(n)$ не може приймати значення менші, ніж $c \cdot g(n)$, тобто

$$f(n) \in \Omega(g(n)), \text{ якщо } \lim_{n\to\infty} \frac{f(n)}{g(n)} \geq c,$$

де $n \geq n_0, g(n) \neq 0$.

Якщо $f(n) = o(g(n))$, то це означає, що функція $g(n)$ зростає набагато швидше, ніж функція $f(n)$, тобто

$$f(n) \in o(g(n)), \text{ якщо } \lim_{n\to\infty} \frac{f(n)}{g(n)} = 0,$$

де $n \geq n_0, g(n) \neq 0$.

Приклад 2.2.4. а) Поліном $f(n) = 4n^5 + 3n^2 + 2n + 3 \in O(n^5)$ оскільки

$$\lim_{n\to\infty} \frac{4n^5+3n^2+2n+3}{n^5} = 4.$$

б) Поліном $f(n) = 4n^5 + 3n^2 + 2n + 3 \in o(n^6)$ і $o(2^n)$ оскільки

$$\lim_{n\to\infty} \frac{4n^5+3n^2+2n+3}{n^6} = 0 \text{ і } \lim_{n\to\infty} \frac{4n^5+3n^2+2n+3}{2^n} = 0.$$

в) Функція $f(n) = 2^n \in o(n!)$ оскільки $\lim_{n\to\infty} \frac{2^n}{n!} = 0$. ♠

Назвемо функцію $f(n)$ *асимптотично додатною*, якщо вона приймає додатні значення при достатньо великих значеннях n.

Деякі основні властивості введеної нотації випливають із наступної теореми.

Теорема 2. *Для довільних асимптотично додатних функцій f(n), g(n), h(n) і l(n) мають місце такі властивості:*

а) $f(n) = O(g(n)) \Leftrightarrow g(n) = \Omega(f(n))$;

б) $f(n) = \Theta(g(n)) \Leftrightarrow f(n) = O(g(n))$ *і* $f(n) = \Omega(g(n))$;

в) *якщо* $f(n) = O(h(n)), g(n) = O(h(n))$, *то* $(f + g)(n) = O(h(n))$;

г) *якщо* $f(n) = O(h(n)), g(n) = O(l(n))$, *то* $(f \cdot g)(n) = O(h(n) \cdot l(n))$;

д) (*рефлексивність*) $f(n) = O(f(n)), f(n) = \Theta(f(n)), f(n) = \Omega(f(n))$;

е) (*симетричність*) $f(n) = \Omega(g(n)) \Leftrightarrow g(n) = \Omega(f(n))$;

є) (*транзитивність*) *якщо*

$f(n) = O(g(n))$ *і* $g(n) = O(h(n))$, *то* $f(n) = O(h(n))$,

$f(n) = \Theta(g(n))$ *і* $g(n) = \Theta(h(n))$, *то* $f(n) = \Theta(h(n))$,

$f(n) = \Omega(g(n))$ *і* $g(n) = \Omega(h(n))$, *то* $f(n) = \Omega(h(n))$,

$f(n) = o(g(n))$ *і* $g(n) = o(h(n))$, *то* $f(n) = o(h(n))$.

Зауважимо, що коли мова йде про ефективність алгоритмів в програмуванні, то найбільший інтерес має клас *O-велике*. А коли мова йде про складність алгоритмів в криптографічних застосуваннях, то тут важливу роль відіграють алгоритми, які мають великі оцінки часової складності.

2.3. Елементи теорії складності

В зв'язку з тим, що строгого означення поняття "алгоритм" не існує, то користуються інтуїтивним означенням алгоритму.

Означення 10. *Алгоритмом називається цілком визначена процедура, яка після отримання вхідних даних генерує вихідні дані в скінченному проміжку часу.*

Зрозуміло, що це означення не є математично строгим, але його достатньо для нашої мети. Вияснимо, що собою являють дані вхідні і вихідні. Для цього скористаємося алфавітом $X = \{0, 1\}$.

Означення 11. *Розміром вхідних (вихідних) даних алгоритму називається довжина слова p в алфавіті X, яке зображує ці дані.*

Пояснимо це поняття прикладами.

Приклад 2.3.1. 1) Натуральне число $n \in \mathcal{N}$ зображується двійковим словом довжини $(1 + \lceil \log\ n \rceil)$, де $\lceil \log\ n \rceil$ означає найменше ціле число, яке не менше $\log\ n$. Далі будемо користатися з запису $\log\ n$ маючи на увазі $\lceil \log\ n \rceil$.

2) Поліном f степеня k, коефіцієнти якого є невід'ємними числами не більшими ніж число n зображується двійковим словом довжини $(k + 1) \cdot \log\ n$.

3) Матриця A з r рядками і s стовпчиками та невід'ємними цілими коефіцієнтами не більшими ніж число n зображується двійковим словом довжини $rs \cdot \log\ n$. ♠

2.3.1 Машинні моделі

Класи складності найбільш природно описуються за допомогою машинних моделей обчислень. Стандартним варіантом таких моделей є машини Тьюрінга (МТ) та їх варіації: детерміновані, недетерміновані, багатострічкові та інші.

Детерміновані машини Тьюрінга. Неформально детермінована машина Тьюрінга (ДМТ) є пристроєм, що має чутливу **головку** G, яка здатна пересуватися вздовж нескінченної в одну (праву) сторону стрічки, розбитої на **клітинки** (див. рис. 2.3.1).

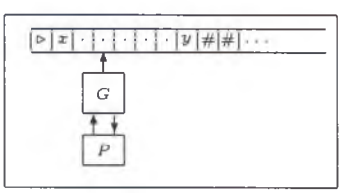


Рис. 2.3.1. Схема ДМТ

Правила руху головки вздовж стрічки визначаються її **програмою** P, яка в теорії ДМТ називається **функцією переходів**. Головка здатна **читати символ зі стрічки, писати символ на стрічку** і **пересуватися** вліво чи вправо на одну клітинку або залишатися в тій самій позиції відповідно до програми даної ДМТ.

Формальне означення ДМТ є таким.

Означення 12. *Детермінованою МТ називається упорядкована четвірка* $M = (K, X, \delta, s_0)$, *де* K *– скінченна множина, елементи якої називаються* **станами ДМТ**, X *– скінченна множина попарно різних елементів, що називається* **алфавітом ДМТ** (*причому* $K \cap X = \emptyset$), *а алфавіт* X *включає символи* $\#$ (*пустий символ*), $\triangleright$ (*початковий символ*) *і пусте слово* e), $\delta : K \times X \to K' \times X \times \{l, r, t\}$ – **функція переходів** *ДМТ, де* $K' = K \cup \{h, \text{“}yes\text{”}, \text{“}no\text{”}\}$ *і* $s_0 \in K$ – **початковий стан ДМТ**.

Символи h, *“yes”*, *“no”* означають відповідно **заключний (або фінальний) стан, стан, що сприймає, і стан, що не сприймає**, а символи l, r, t означають напрямок руху головки ДМТ відповідно вліво, вправо і незмінність позиції. Нехай X^* означає множину всіх скінченних слів у алфавіті X.

Робота ДМТ відбувається таким чином. Деяке слово $q = \triangleright\, p \in X$ записано на стрічці ДМТ. У початковий момент часу ДМТ перебуває в початковому стані s_0 і оглядає перший символ слова q – символ $\triangleright$. При цьому слово $p \in X^*$ називається **вхідним словом** або **вхідними даними ДМТ**. Виходячи зі своєї початкової комбінації, ДМТ виконує крок обчислень таким чином. Для кожної комбінації поточного стану $k \in K$ і символа $x \in X$ функція δ визначає трійку (k', y, d), де k' – черговий стан, $y \in X$ – символ, який записується на стрічці замість символа x, а $d \in \{l, r, t\}$ визначає напрямок руху головки ДМТ. У випадку, коли робота ДМТ тільки починається, завжди $\delta(s_0, \triangleright) = (k', \triangleright, r)$, тобто символ $\triangleright$ завжди приводить до руху головки ДМТ вправо і на його місце не можна нічого записувати. Ця обставина дає можливість розглядати ДМТ зі стрічкою, необмеженою в одну (праву) сторону, оскільки вихід за початковий символ $\triangleright$ у такій ситуації неможливий.

Попри те, що головка не може вийти за лівий кінець вхідного слова p, вона може вийти за його правий кінець. У цьому разі говорять, що головка читає пустий символ $\#$, на місце якого може бути записаний якийсь символ з алфавіту X. Таким чином, послідовність символів на стрічці може зростати, а це необхідно для того, щоб виконувати довільні обчислення за допомогою ДМТ.

Зупинка ДМТ настає тоді, коли вона в процесі обчислень досягає одного з трьох станів h, *“yes”*, *“no”*. При цьому, якщо ДМТ зупинилась

– у стані *“yes”*, то говоримо, що ДМТ **сприймає** або **розпізнає вхідне слово** $p \in X$;

– у стані *“no”*, то говоримо, що ДМТ **не сприймає** або **не розпізнає вхідне слово** $p \in X$;

– у стані h, то результатом роботи ДМТ вважається слово, записане на стрічці ДМТ в момент її зупинки.

Оскільки обчислення у разі зупинки ДМТ відбуваються в скінченному часі, то на стрічці буде записане слово скінченної довжини, яке починається символом $\triangleright$ і

закінчується символом, який передує першому пустому символу, якщо дивитися символи слова зліва направо (це слово може бути, зокрема, і пустим). Якщо слово q є результатом роботи ДМТ на вхідному слові p, то пишемо $M(p) = q$.

Може трапитися й таке, що ДМТ ніколи не досягне якогось зі станів h, "yes", "no" і тому її робота триватиме нескінченно. У цьому випадку вважається, що результат роботи ДМТ на вхідному слові p **невизначений** або, що ДМТ **незастосовна до вхідного слова** p.

Як бачимо, наведене означення ДМТ відрізняється від загальноприйнятого, оскільки в загальноприйнятому означенні ДМТ її стрічка нескінченна в обидві сторони.

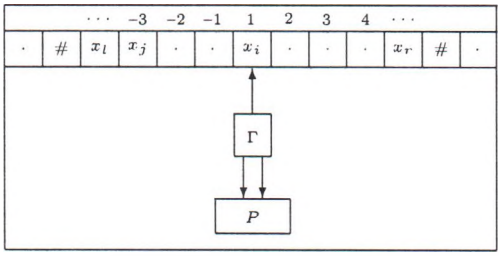


Рис. 2.3.2. ДМТ з нескінченною в обидві сторони стрічкою

Але це не впливає на обчислювальну можливість МТ, оскільки має місце

Теорема 3. *Для довільної ДМТ M з нескінченною в обидві сторони стрічкою існує еквівалентна їй ДМТ M' з нескінченною в одну (праву) сторону стрічкою [9].*

Завдяки цій теоремі далі будемо користуватися лише поняттям ДМТ з нескінченною в одну (праву) сторону стрічкою, оскільки обчислювальна потужність ДМТ при цьому не змінюється.

Операції, що виконуються ДМТ в процесі обчислень, можна визначити формально за допомогою поняття **конфігурації ДМТ**.

Означення 13. *Конфігурацією ДМТ називається трійка (s, p, q), де $s \in K$, а слова p і q із X^* такі, що p є словом, яке записане на стрічці ДМТ зліва від клітинки, яку оглядає головка, включаючи і символ, записаний в цій клітинці, а q – слово, записане на стрічці справа від клітинки, яку оглядає головка ДМТ.*

Означення 14. *Нехай M – деяка ДМТ. Говоримо, що ДМТ M з конфігурації (s, p, q) безпосередньо досягає конфігурації (s', p', q') (позначення $(s, p, q) \xrightarrow{M} (s', p', q')$), якщо виконуються такі умови:*

а) *$\delta(s, x) = (s', y, d)$, якщо x останній символ слова p;*

б) *якщо $d = l$, то p' є словом, що отримане зі слова p шляхом викреслювання символа x, а $q' = yq$;*

в) *якщо $d = r$, то $p' = py$, а q' є словом, що отримане зі слова q шляхом викреслювання його першого символа;*

г) *якщо $d = t$, то p' є словом, у якого останній символ x замінений на символ y, а $q' = q$.*

Нехай M^* означає транзитивне замикання відношення безпосередньої досяжності для конфігурацій ДМТ M (позначення $(s,p,q) \xrightarrow{M^*} (s',p',q')$). Відношення M^* називається просто відношенням досяжності для конфігурацій і означає, що існує таке натуральне число $n \in N$, для якого справедливі такі переходи $(s,p,q) \xrightarrow{M^n} (s',p',q')$, тобто для всіх $i = 1,2,\ldots,n$ маємо $s = s_1$, $s_{n+1} = s'$, $p_1 = p$, $q_{n+1} = q'$ і $(s_i,p_i,q_i) \xrightarrow{M} (s_{i+1},p_{i+1},q_{i+1})$. У цьому випадку говорять також, що ДМТ із конфігурації (s,p,q) досягає конфігурації (s',p',q') або переходить до неї за n кроків.

МТ як алгоритми. МТ є природними моделями для ~~розв'язання проблем на словах~~, а саме **обчислення функцій на словах**, а також **розпізнавання** і **розв'язуваності мов**.

Означення 15. *Нехай $L \subseteq (X \setminus \{\#\})^*$ - деяка мова в алфавіті $X' = X \setminus \{\#\}$, а M - МТ така, що для довільного слова $p \in (X \setminus \{\#\})^*$ отримуємо $M(p) =$ "yes", якщо $p \in L$, і $M(p) =$ "no", якщо $p \notin L$. У цьому випадку говорять, що ДМТ M* **розв'язує мову** *L.*

Мова L називається **рекурсивною**, *якщо існує ДМТ, яка розв'язує цю мову.*

Говоримо, що ДМТ M тільки **розпізнає мову** *L, якщо для довільного слова $p \in (X \setminus \{\#\})^*$ маємо $M(p) =$ "yes" тоді і тільки тоді, коли $p \in L$ і M незастосовна до p, якщо $p \notin L$.*

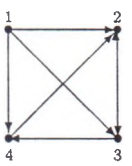
Мова L називається **рекурсивно переліковною**, *якщо вона розпізнається деякою ДМТ.*

МТ застосовуються не тільки для розв'язуваності і розпізнавання мов, а й для обчислення значень словарних функцій.

Означення 16. *Функція $f : (F(X) \setminus \{\#\}) \to F(X)$, де X - деякий алфавіт, називається* **словарною функцією**.

Нехай M - деяка ДМТ в алфавіті X. Говорять, що словарна функція f обчислюється ДМТ M, якщо для довільного слова $p \in (F(X) \setminus \{\#\})$ має місце рівність $M(p) = f(p)$. Якщо така ДМТ існує, то функція f називається **частково рекурсивною**.

Представлення проблем за допомогою слів. Слова в алфавіті $X \setminus \{\triangleright, \#\} = \{0,1\}$, про що йшлося вище, служать для представлення проблем. За такого підходу має бути очевидним, що представлення проблеми у вигляді слова в алфавіті $\{0,1\}$ є достатньо загальним: довільний скінченний об'єкт може бути представлений словом скінченної довжини в цьому алфавіті. Наприклад, такий об'єкт, як скінченний граф, може бути записаний у вигляді слова скінченної довжини в алфавіті $\{0,1\}$. Справді, використовуючи матрицю інцидентності, можна наведений нижче граф подати таким словом у цьому алфавіті:



"(1,10),(1,11),(1,100),(10,11),
(11,100),(100,10)"
списки інцидентності
"(0111, 0010, 0001, 0100)"
матриця інцидентності

Як випливає з цього простого прикладу, цілі числа, скінченні множини, скінченні графи і т. п. можна зображувати різними способами, і такі представлення можуть відрізнятися одне від одного як формою, так і довжиною. Але відома важлива властивість таких зображень:

Всі допустимі зображення-кодування є поліноміально еквівалентними.

Слова "поліноміально еквівалентні" означають, що коли A і B – два сенсовні представлення проблеми P і представлення A у вигляді слова має довжину n, то представлення B тієї самої проблеми P у вигляді слова має найбільшу довжину $p(n)$, де $p(n)$ – деякий поліном. Наприклад, представлення скінченного графа без ізольованих вершин у вигляді матриці інцидентності вимагає найбільшої довжини слова, яка є квадратом довжини представлення цього графа у вигляді списків інцидентності.

Зауважимо, що теорія складності обчислень не залежить від проблеми представлення і її можна розв'язувати незалежно від слів і мов. Але розумний вибір представлення робить результати теорії складності адекватними реальним проблемам і практичній обчислювальності.

2.3.2. Багатострічкові МТ

Для означення часу і пам'яті, необхідних для оцінки складності даної проблеми, потрібне певне узагальнення ДМТ, а саме, багатострічкові МТ, тобто ДМТ з декількома стрічками. Це узагальнення, як буде показано далі, можна змоделювати за допомогою звичайної однострічкової ДМТ. Отже, прийняття більшої кількості стрічок в моделі ДМТ не виводить нас за клас МТ.

Означення 17. *ДМТ із k стрічками ($k \geq 1$) називається четвірка $M = (K, X, \delta, s_0)$, де K і X ті ж, що і в означенні звичайної однострічкової ДМТ, а функція переходів δ, яка називається програмою, визначає наступний стан таким чином:*

$$\delta : K \times X^k \to (K \cup \{h, \text{"так"}, \text{"ні"}\}) \times (X \times \{l, r, t\})^k,$$

де $\delta(s, y_1, \ldots, y_k) = (s', z_1, d_1, \ldots, z_k, d_k)$ означає, що коли ДМТ в деякий момент перебуває в стані s, головка на першій стрічці оглядає символ y_1 і т. д., головка на k-й стрічці оглядає символ y_k, то в наступний момент ДМТ перебуватиме в стані s', головка на першій стрічці запише символ z_1 замість символа y_1 і перейде або залишиться на місці залежно від значення d_1 і т. д., головка на k-й стрічці запише символ z_k замість символа y_k і перейде або залишиться в тій самій позиції, залежно від значення d_k.

На місце символа $\triangleright$ не дозволяється нічого записувати, тобто це означає, що коли $y_i = \triangleright$, то $d_i = r$. На початку на всіх стрічках записаний символ $\triangleright$, а на першій стрічці записано ще й вхідне слово p (тобто $q = \triangleright\, p$).

Результат роботи багатострічкової ДМТ визначається так само, як і для звичайної ДМТ, з тією лише відмінністю, що результат обчислень словарної функції після зупинки ДМТ записується на останній k-й стрічці.

Приклад 2.3.1. Паліндроми можна розв'язувати ефективніше за допомогою двострічкової ДМТ, порівняно з однострічковою ДМТ. Двострічкова ДМТ починає свою роботу з копіювання вхідного слова з першої стрічки на другу. Після цього ДМТ встановлює головку першої стрічки на перший символ вхідного слова, а головку другої стрічки – на останній символ скопійованого слова. А далі робота ДМТ зводиться до руху головок у протилежних напрямках і порівняння на кожному кроці символів, що оглядаються головками. У разі рівності цих символів символ на другій стрічці замінюється пустим (тобто цей символ просто стирається на другій стрічці).

Таблицю функції переходів цієї ДМТ наведено нижче.

$k \in K$	$x \in X$	$y \in X$	$\delta(k,x,y)$
s_0	0	#	$(s_0, 0, r, 0, r)$
s_0	1	#	$(s_0, 1, r, 1, r)$
s_0	▷	▷	$(s_0, \triangleright, r, \triangleright, r)$
s_0	#	#	$(s, \#, l, \#, t)$
s	0	#	$(s, 0, l, \#, t)$
s	1	#	$(s, 1, l, \#, t)$
s	▷	#	$(s_1, \triangleright, r, \#, l)$
s_1	0	0	$(s_1, 0, r, \#, l)$
s_1	1	1	$(s_1, 1, r, \#, l)$
s_1	0	1	$(\text{“}no\text{”}, 0, t, 1, t)$
s_1	1	0	$(\text{“}no\text{”}, 1, t, 0, t)$
s_1	#	▷	$(\text{“}yes\text{”}, \#, t, \triangleright, r)$

♠

Означення 18. *Конфігурацією k-стрічкової ДМТ M називається кортеж вигляду $(s, p_1, q_1, \ldots, p_k, q_k)$, де s – поточний стан ДМТ, p_i, q_i – слова, що записані на i-й стрічці $(1 \le i \le k)$, головка якої оглядає останній символ слова p_i.*

Говоримо, що ДМТ M з конфігурації $(s, p_1, q_1, \ldots, p_k, q_k)$ безпосередньо досягає конфігурації $(s', p'_1, q'_1, \ldots, p'_k, q'_k)$ (позначення $(s, p_1, q_1, \ldots, p_k, q_k) \xrightarrow{M} (s', p'_1, q'_1, \ldots, p'_k, q'_k)$), якщо виконуються такі умови. Нехай $p_i = p_{i-1}y_i$, $q_i = xq_{i-1}$ для $i = 1, 2, \ldots, k$ і нехай $\delta(s, y_1, \ldots, y_k) = (s', z_1, d_1, \ldots, z_k, d_k)$. Тоді для кожного $i = 1, 2, \ldots, k$

- *якщо $d_i = r$, то $p'_i = p_{i-1}z_ix$, а $q'_i = q_{i-1}$;*
- *якщо $d_i = l$, то $p'_i = p_{i-1}$, а $q'_i = z_ixq_{i-1}$;*
- *якщо $d_i = t$, то $p'_i = p_{i-1}z_i$, а $q'_i = q_i$.*

Тобто, для кожної стрічки мають виконуватися умови переходу від однієї конфігурації до іншої як для ДМТ з однією стрічкою.

Транзитивне замикання відношення безпосередньої досяжності для конфігурацій k-стрічкової ДМТ називається **відношенням досяжності** *для конфігурацій (позначення $(s, p_1, q_1, \ldots, p_k, q_k) \xrightarrow{M^*} (s', p'_1, q'_1, \ldots, p'_k, q'_k)$). Це означає, що існує таке число $n \in N$, що*

$$(s, p_1, q_1, \ldots, p_k, q_k) \xrightarrow{M^n} (s', p'_1, q'_1, \ldots, p'_k, q'_k),$$

і в цьому випадку будемо говорити, що друга конфігурація досяжна з першої конфігурації за n кроків.

k-стрічкова ДМТ починає обчислення на вхідному слові p перебуваючи в конфігурації

$$(s_0, \triangleright, p, \triangleright, e, \ldots, \triangleright, e),$$

де e – пусте слово. Це означає, що вхідне слово p записується на першій стрічці, а на решті стрічок записаний тільки початковий символ $\triangleright$. Якщо

$$(s_0, \triangleright, p, \triangleright, e, \ldots, \triangleright, e) \xrightarrow{M^*} (\text{“}yes\text{”}, p_1, q_1, \ldots, p_k, q_k)$$

для деяких слів $p_1, q_1, \ldots, p_k, q_k$, то говоримо, що $M(p) =$ “yes”. Якщо $(s_0, \triangleright, p, \triangleright, e, \ldots, \triangleright, e) \xrightarrow{M^*} (\text{“}no\text{”}, p_1, q_1, \ldots, p_k, q_k)$ для деяких слів $p_1, q_1, \ldots, p_k, q_k$, то говоримо, що $M(p) =$ “no”. Нарешті, якщо ДМТ зупиняється в конфігурації $(h, p_1, q_1, \ldots, p_k, q_k)$, то $M(p) = = p'_k q_k$, де p'_k збігається зі словом p_k, взятим без першого символа $\triangleright$, а q_k – слово, яке записане без пустих символів. Отже, у разі зупинки k-стрічкової ДМТ в стані h результат записується на останній k-й стрічці. Завдяки цьому звичайна однострічкова ДМТ є окремим випадком k-стрічкової ДМТ за $k = 1$. Крім того, це дає змогу визначити поняття обчислюваної словарної (рекурсивної) функції, розв'язуваності й розпізнавання мов аналогічно тому, як це визначалось для однострічкових ДМТ.

2.3.3. Класи $P, PSPACE, L$ і нижче

Далі нам будуть потрібні не довільні k-стрічкові ДМТ, а деякі спеціальні. Зокрема, для побудови класів складності використовуються тристрічкові ДМТ. Точніше, ДМТ має 3 стрічки, серед яких

– на першій стрічці, яка називається **вхідною**, записується вхідне слово і з цієї стрічки дозволяється лише читати і не дозволяється нічого писати;

– друга стрічка, яка називається **робочою** або **внутрішньою**, така, що на неї можна писати і з неї можна читати;

– третя стрічка, яка називається **вихідною**, така, що на неї дозволяється тільки писати і не дозволяється нічого з неї читати.

Розв'язання проблеми такою машиною виконується так. На вхідній стрічці записується вхідне слово в алфавіті $X = \{0, 1, \triangleright, \#\}$ починаючи з найлівішої клітинки (тобто з першої клітинки стрічки), а в решту клітинок стрічки записаний символ # – символ пустої клітинки. Машина працює з цим словом відповідно до своєї програми (функції переходів), записує необхідні символи на вихідну і робочі стрічки. Запис на вихідну стрічку теж виконується з самої лівої клітинки, а в решту клітинок стрічки записаний символ #. Якщо в деякий момент машина зупиняється з певною відповіддю, то вона записує цю відповідь – “yes” або “no” – на вихідну стрічку.

В теорії складності обчислень принциповий інтерес мають два ресурси для базової моделі обчислень – *час* і *пам'ять*.

Означення 19. *Якщо для тристрічкової ДМТ і вхідного слова p справедливе відношення*

$$(s_0, \triangleright, p, \triangleright, e, \triangleright, e) \xrightarrow{M^t} (H, p_1, q_1, p_2, q_2, p_3, q_3),$$

де $H \in \{h, \text{“}yes\text{”}, \text{“}no\text{”}\}$, то число t називається **часом обчислень** *або* **часовою складністю** *розв'язання проблеми. Інакше кажучи, часовою складністю розв'яза-*

ння проблеми називається число кроків тристрічкової ДМТ, які вона виконала до зупинки.

Пам'яттю *або* **простором обчислень** *називається число клітинок робочої стрічки, які використала ДМТ під час обчислень.*

Зауважимо, що для ДМТ пам'ять, яка використовувалась під час обчислень, може бути набагато меншою, ніж розмір вхідного слова, а час обчислень має бути принаймні не меншим від довжини вхідного слова. Це пов'язано з тим, що відповідь часто залежить тільки від деякого початкового підслова вхідного слова, а не від усього слова.

Означення 20. *Нехай $f : N \to N$ деяка функція. Говоримо, що ДМТ M* **працює в часі** *$f(n)$, якщо для довільного вхідного слова p час обчислень, необхідний ДМТ M для розв'язання проблеми p, дорівнює $f(l(p))$, де $l(p) = n$ – довжина слова p. Функція f називається* **часовим обмеженням** *ДМТ M.*

Тепер можна ввести поняття класу складності, використовуючи тристрічкові ДМТ і розв'язувані ними мови.

Означення 21. *Нехай $L \subseteq (X \cup \{\#\})^*$ – мова. Говоримо, що мова L належить* **класу** $\mathbf{TIME}(f(n))$, *якщо існує тристрічкова ДМТ, яка розв'язує мову L в часі $f(n)$. Множину мов* $\mathbf{TIME}(f(n))$ *називають* **класом часової складності**.

Якщо мова L розв'язується тристрічковою ДМТ в поліноміальному часі, тобто $f(n) = n^k$, де k – стала величина, то клас таких мов за всіма k складає **поліноміальний клас часової складності** *і позначається* $\mathbf{P}$. *Отже,*

$$\mathbf{P} = \mathbf{TIME}(n^k) = \bigcup_{j>1} \mathbf{TIME}(n^j).$$

Говоримо, що мова L належить до класу $\mathbf{SPACE}(f(n))$, *якщо існує тристрічкова ДМТ, яка розв'язує мову L і використовує не більше $f(n)$ клітинок робочої стрічки. Множину мов* $\mathbf{SPACE}(f(n))$ *називають* **класом складності по пам'яті**.

Якщо мова L розв'язується тристрічковою ДМТ при поліноміальній пам'яті, тобто $f(n) = n^k$, де k – стала величина, то клас таких мов за всіма k складає **поліноміальний клас складності по пам'яті** *і позначається* $\mathbf{PSPACE}$, *тобто*

$$\mathbf{PSPACE} = \mathbf{SPACE}(n^k) = \bigcup_{j>1} \mathbf{SPACE}(n^j).$$

Говоримо, що мова L належить класу $\mathbf{SPACE}(\log n)$, *якщо існує тристрічкова ДМТ, яка розв'язує мову L і використовує не більше $\log n$ клітинок робочої стрічки. Множину мов* $\mathbf{SPACE}(\log n)$ *називають* **класом логарифмічної складності по пам'яті** *і позначають через* $\mathbf{L}$ *(або іще* $\mathbf{LOGSPACE}$*), тобто*

$$\mathbf{L} = \mathbf{SPACE}(\log n).$$

Зауважимо, що визначені вище класи складності $\mathbf{P}$, $\mathbf{PSPACE}$, $\mathbf{L}$ є найважливішими і найпопулярнішими класами в теорії складності обчислень. Розглянемо приклади.

Приклад 2.3.2. Неважко показати, що ДМТ із прикладу 2.3.1 розпізнає паліндроми в часі, пропорційному величині $f(n) = n^2$. Робота цієї ДМТ виконується в два етапи. На першому етапі за $2n+1$ кроків виконується порівняння першого і останнього символів. Потім повторюється та сама робота над словом довжини $n-2$, далі – над словом довжини $n-4$ і т. д. Отже, загальне число кроків ДМТ у найгіршому випадку буде $f(n) = \frac{(n+1)(n+2)}{2}$. Звідси випливає, що мова паліндромів належить до класу $\mathbf{TIME}(\frac{(n+1)(n+2)}{2}) = O(n^2)$. Слід зазначити, що одержана оцінка є найбільш песимістичною, тому що для слова 01^n ця ДМТ за $n+3$ кроків визначає, що дане слово не є паліндромом. Але, обчислюючи $f(n)$, необхідно брати до уваги найгірший із можливих варіантів вхідних даних.

Нарешті зазначимо, що двострічкова ДМТ з прикладу 2.3.1 розв'язує мову паліндромів у часі $f(n) = 3n+3 = O(n)$, звідки випливає, що ця мова належить класу складності $\mathbf{TIME}(3n+3)$. ♠

Аналогічно до означення класів **P, PSPACE, L** визначаються і нижченаведені класи складності, які часто трапляються у застосуваннях. Ці класи лежать в ієрархії нижче від класу **P** і їх означення таке:

– **константа:** існує така константа k, що мова L розв'язується ДМТ за k кроків. Клас всіх таких мов позначається $\mathbf{TIME}(k)$;

– **логарифм:** мова L розв'язується ДМТ за $\log n$ кроків. Клас усіх таких мов позначається $\mathbf{TIME}(\log n)$;

– **полілогарифм:** існує така константа k, що мова L розв'язується ДМТ за $\log^k n$ кроків. Клас усіх таких мов позначається $\mathbf{TIME}(\log^k n)$;

– **лінійна:** мова L розв'язується ДМТ за n кроків. Клас усіх таких мов позначається $\mathbf{TIME}(n)$.

Аналогічно визначаються і класи складності по пам'яті.

Наведена коласифікація грунтується на наступному твердженні.

Теорема 4. *Для довільної k-стрічкової ДМТ M, яка працює в часі $f(n)$, існує однострічкова ДМТ M', що працює в часі $O(f(n)^2)$, така, що для довільного слова $p \in F(X)$ має місце рівність $M(p) = M'(p)$ [9].*

З цієї теореми випливає, що збільшення числа стрічок не збільшує обчислювальної потужності багатострічкових ДМТ порівняно з однострічковими. Єдине, до чого приводить збільшення кількості стрічок, так це поліноміальне підвищення ефективності такої ДМТ.

2.3.4. Недетерміновані МТ. Класи $NP, NSPACE$ і EXP

Недетермінована МТ визначається аналогічно ДМТ, тільки відношення переходів у такій МТ ~~є відношенням~~, а не функцією. Це означає, що множина можливих обчислень може розгалужуватися, як дерево. ~~Наведемо формальні означення~~.

Означення 22. *Недетермінованою МТ (НДМТ) називається четвірка $M = (K, X, \Delta, s_0)$, де K, X, s_0 – ті самі об'єкти, що і в ДМТ, а Δ – відношення переходів, яке визначає декілька можливих станів, у які МТ може перейти. Це значить, що НДМТ має можливість вибору між кількома діями і Δ вже не є функцією, тобто*

$$\Delta \subset (K \times X) \times ((K \cup \{h, \text{“}yes\text{”}, \text{“}no\text{”}\}) \times X \times \{l, r, t\}).$$

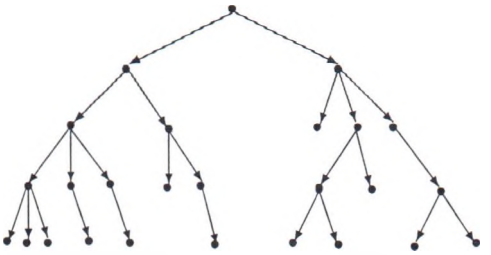


Рис. 2.3.3. Недетерміновані обчислення

Слід зауважити, що загальний час обчислень НДМТ може бути набагато більшим, ніж час, який вважається часом обчислень НДМТ (це легко помітити з наведеного вище рисунка).

Множина мов, що розв'язуються НДМТ, є новим важливим класом складності, і позначається цей клас $\mathbf{NTIME}(f(n))$. Важливим класом складності є клас **NP**, який означається як теоретико-множинне об'єднання класів $\mathbf{NTIME}(n^k)$, де k – деяке стале число, тобто

$$\mathbf{NP} = \bigcup_{k>1} \mathbf{NTIME}(n^k).$$

Зауважимо, що клас **P** є підкласом класу **NP**, оскільки клас ДМТ є підкласом НДМТ. Іншими важливими класами складності є класи

$$\mathbf{NPSPACE} = \mathbf{NSPACE}(n^k) \text{ і } \mathbf{EXP} = \mathbf{TIME}(2^{n^k}).$$

Для класів складності за пам'яттю можна розглядати класи нижче від лінійної. Двома такими класами є вже відомий нам клас $\mathbf{L} = \mathbf{SPACE}(\log n)$ і клас $\mathbf{NL}$=$\mathbf{NSPACE}(\log n)$. Цими класами далеко не вичерпуються всі теоретично можливі класи складності [31].

Між ДМТ і НДМТ існує зв'язок, який виражається таким твердженням [31].

Теорема 5. *Нехай деяка НДМТ NM розв'язує мову L у часі f(n), тоді існує тристрічкова ДМТ M, яка розв'язує мову L у часі $O(c^{f(n)})$, де c – константа, що залежить від NM, тобто*

$$\mathbf{NTIME} \subseteq \bigcup_{c>1} \mathbf{TIME}(c^{f(n)}).$$

Тепер уточнимо поняття обмеженості за пам'яттю для тристрічкової НДМТ $NM = (K, X, \Delta, s_0)$.

Означення 25. *Говоримо, що НДМТ NM розв'язує мову L у пам'яті $f(n)$, якщо NM розв'язує мову L і для довільного $p \in (X\setminus\{\#\})^*$, для якого $(s_0, \triangleright, p, \triangleright, e, \triangleright, e) \xrightarrow{NM^*} (s, p_1, q_1, p_2, q_2, p_3, q_3)$, має місце нерівність $l(p_2q_2) \le f(l(p))$,* де $n = l(p)$.

Це означає, що НДМТ NM під час обчислень жодного разу не потребує клітинок на її робочій стрічці більше, ніж значення функції f на довжині вхідного слова p. Зауважимо, що в даному означенні навіть не вимагається, щоб усі обчислення НДМТ

були скінченними. Та все ж, для означення деяких типів проблем і алгоритмів за допомогою НДМТ будемо припускаються того, що її дерево досяжних конфігурацій є скінченним і включає всі фінальні конфігурації, тобто конфігурації, в яких НДМТ зупиняється, причому глибина цих конфігурацій така ж, як і глибина самого дерева. У цьому разі простором-пам'яттю, що використовується, є максимальне число клітинок на робочій стрічці, що відвідуються, за всіма конфігураціями цього дерева.

2.3.5 Редукція і повнота

Для того, щоб можна було порівнювати проблеми за складністю їх розв'язання, потрібне поняття, яке дає змогу точно окреслити ситуацію, коли дві проблеми мають однаковий ступінь складності. Таким поняттям в теорії складності обчислень є поняття **редукції**. Неформально редукція проблеми B до проблеми A означає існування функції R, в результаті якої для довільного окремого випадку проблеми B є еквівалентний йому окремий випадок $R(p)$ проблеми A. Під еквівалентністю двох окремих випадків розуміють те, що відповідь на питання, чи є $R(p)$ позитивним прикладом проблеми A, буде одночасно відповіддю на питання, чи є p позитивним прикладом проблеми B. Інакше кажучи, розв'язати проблему B для прикладу p це те саме, що розв'язати проблему A для прикладу $R(p)$.

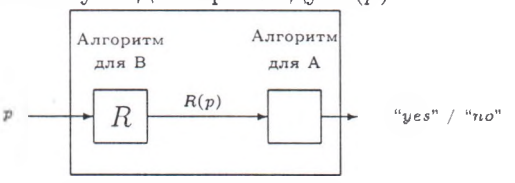


Рис. 2.3.4. Редукція B до A

Якщо має місце ситуація, показана на рисунку, то говоримо, що проблема A є не менш важкою, ніж проблема B за одного застереження. Це застереження стосується редукції R: редукція R не повинна бути складною в обчислювальному плані. Якщо не зробити жодних обмежень відносно цієї складності, то можна дійти до абсурдних наслідків. Формальне означення редукції має вигляд.

Означення 26. *Говоримо, що мова* L_1 **редукується до мови** L_2, *якщо існує словарна функція* R (*функція, визначена на словах і зі значеннями у множині слів*), *яка обчислюється за допомогою ДМТ в логарифмічній пам'яті* (*тобто* $O(\log n)$), *така, що для кожного слова* p *має місце* $p \in L_1$ *тоді і тільки тоді, коли* $R(p) \in L_2$. *При цьому функція* R *називається* **редукцією** L_1 *до* L_2.

Позначимо $L \leq_R L'$ той факт, що мова L редукується до мови L' за допомогою редукції R.

Звідси безпосередньо випливає, що редукції є поліноміальними алгоритмами, і це стверджує наступне твердження.

Теорема 6. *а) Якщо* R *є редукцією, яка обчислюється за допомогою деякої ДМТ* M, *то для кожного вхідного слова* p *машина* M *закінчує свою роботу через поліноміальне число кроків [31].*

б) Нехай L_1, L_2 L_3 алгоритмічно розв'язувані проблеми. Тоді
б1) (транзитивність) якщо $L_1 \leq_{R_1} L_2$ і $L_2 \leq_{R_2} L_3$, то $L_1 \leq_R L_3$, де $R = R_1 * R_2$;
б2) якщо $L_1 \leq_R L_2$ і $L_2 \in P$, то $L_1 \in P$.

Означення 27. *Проблема L називається NP-повною, якщо*
(1) $L \in NP$ і (2) $\forall L_1 \in NP(L_1 \leq_R L)$.

Клас NP-повних проблем позначається NPC.

Приклад NPC проблем. а) **(Підмножина суми)** *Вхід:* множина невід'ємних цілих чисел $\{a_1, a_2, \ldots, a_n\}$ і натуральне число $S > 0$.

Проблема: існує чи ні підмножина чисел $\{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\}$, така що $\sum_{j=1}^{k} a_{i_j} = S$?

Відомо, що ця проблема в загальному випадку належить до класу NPC.

б) **(Дискретний логарифм)** *Вхід:* рівняння $y = g^x \bmod p$, де p просте число, а g – елемент скінченного поля $\mathcal{F}_q$.

Проблема: обчислити x, якщо відомі числа y, g, p. Число x називається **дискретним логарифмом**.

Відомо, що ця проблема належить класу NPC і складність розв'язання цієї проблеми еквівалентна складності розв'язання проблеми факторизації натурального числа.

Найкращий алгоритм розв'язання проблеми факторизації натуральних чисел має складність $e^{((\ln\ p)^{1/3}(\ln(\ln\ p))^{2/3}(c+o(1))}$. Для великих чисел це дуже складна задача. Дійсно, нехай комп'ютер обчислює добуток двох чисел з 90 знаками в часі 10^{-14} (для сучасних комп'ютерів цей час нереальний). Нехай модуль має від 50 до 100 знаків, тоді час, потрібний комп'ютеру для факторизації, дорівнює $T = 2 \cdot \log\ p \cdot 10^{-14} = 6 \cdot 10^{-12}$ секунд. А обчислення дискретного логарифму потребує $10^{45} \cdot 10^{-14} = 10^{31}$ сек. $> 10^{22}$ років. ♠

Теорема 7. *Нехай L_1 і L_2 – алгоритмічно розв'язувані проблеми. Тоді якщо $L_1 \in NP$, $L_2 \in NPC$ і $L_2 \leq_R L_1$, то $L_1 \in NPC$.*

Звідси випливає, що для доведення того, що проблема L_1 є NP-повною потрібно
1) довести, що $L_1 \in NP$;
2) знайти відому проблему $L_2 \in NPC$;
3) довести, що $L_2 \leq_R L_1$.

Означення 28. *Алгоритмічно розв'язувана проблема L називається NP-важкою, якщо існує проблема $L_1 \in NPC$ і $L_1 \leq_R L$.*

Описані вище класи складності називають класами складності в моделі Тьюрінга. В практичних застосуваннях користуються іншою моделлю складності, яку називають *арифметичною*. В моделі Тьюрінга при знаходженні складності враховується довжина зображення вхідних даних, а в арифметичній моделі лише кількість елементарних операцій, які виконуються над цими даними.

Означення 29. *В арифметичній моделі часовою складністю алгоритму на вхідних даних p називається кількість елементарних операцій, які виконуються алгоритмом для обчислення значень вихідних даних.*

"Елементарні операції" трактуються різними способами: операції на бітах, порівняння слів або бітів, елементарні інструкції процесора, додавання або множення бінарних чисел тощо.

Алгоритм називається **строго поліноміальним**, якщо він має часову поліноміальну складність в арифметичній моделі і логарифмічну складність за пам'яттю в моделі Тьюрінга (або, що те саме, поліноміальну часову складність в моделі Тюрінга).

2.3.6. Односторонні функції

З вищенаведених означень випливає, що клас скланості P включається в клас складності NP, а знаменитою відкритою проблемою є питання: збігаються ці класи, чи ні? (проблема тисячоліття). Загальне переконання полягає в тому, що $P \neq NP$ і тому NP-повні проблеми розглядаються як важкі для розв'язання. Але в криптографічних застосуваннях теорії складності необхідно дати відповідь на питання: чи є гіпотеза $P \neq NP$ необхідною і достатньою умовою для існування стійких криптографічних систем?

Необхідність майже очевидна, а достатність не завжди має місце. Справа в тому, що навіть коли $P \neq NP$, то довільна NP-повна задача може виявитися важкою лише для деякої відносно невеликої підмножини вхідних даних, в той час як на всіх інших вхідних даних вона має поліноміальну складність. Це пов'язано з тим, що у означення класу складності NP покладена міра складності "в найгіршому випадку". Таким чином, стає зрозумілим, що для криптографічної стійкості необхідно мати сильнішу умову, ніж $P \neq NP$. А саме, припущення про існування односторонніх функцій.

Про такого типу функції говорилося в пункті а) властивості 1. Користуючись вищенаведеними поняттями теорії складності обчислень, тепер можна уточнити поняття односторонньої функції.

Означення 30. *Функція f(x) називається односторонньою, якщо*

а) для кожного значення аргумента x складність обчислення значення y=f(x) належить класу P, а складність обчислення значення $x = f^{-1}(y)$ належить класу NP і це значення практично неможливо обчислити за розумний проміжок часу;

б) не існує пари значень x, x' таких, що $x \neq x'$ і $f(x) = f(x')$.

В криптографічних системах з відкритим ключем використовуються функції, які ввели Діффі і Хеллман в 1976 році і які назвали *односторонніми функціями з секретом.*

Означення 31. *Одностороння функція з секретом $f_k(x) = y$ називається функція, складність обчислення якої для всіх значень x належить до класу P, але обчислення $x = f_k^{-1}(y)$ майже для всіх значень y неможливе в розумному проміжку часу. Але, якщо скористатися секретною інформацією k, то для всіх значень y обчислення значення x такого, що $f_k(x) = y$ належить класу P.*

Це поняття є основним в криптографії з відкритим ключем. Необхідною умовою існування односторонніх функцій служить гіпотеза $P \neq NP$. А питання існування одно-

сторонніх функцій в загальному випадку залищається відкритим в зв'яку з тим, що нинішній стан наших знань не дозволяє строго довести, що односторонні функції взагалі існують. Не дивлячись на це, існують кандидати серед функцій, ефективне обчислення значень яких нам відоме, тоді як жодні ефективні алгоритми обчислення значень обернених функцій невідомі.

Першим кандидатом на односторонню функцію є множення в кільці цілих чисел $\mathcal{Z}$. Дійсно, множення навіть дуже великих цілих чисел відносно просте, в той час як самий потужний комп'ютер з найкращим в його розпорядженні алгоритмом не в стані швидко (в поліноміальному часі) розкласти на множники тисячозначне ціле число, яке є добутком двох приблизно рівних простих чисел.

Другим важливим кандидатом на односторонню функцію є модульне піднесення до степеня або логарифмування в кільці лишків за модулем n – Z_n. Піднесення до степеня обчислюється в цьому кільці швидко, а знаходження значення оберненої функції (логарифма) практично неможливе. Розглянемо приклади.

Приклади односторонніх функцій. 1) Нехай $X = \{1, 2, \ldots, 16\}$ і $f(x) = r_x$ для кожного $x \in X$, де r_x – остача від ділення 3^x на 17. Тоді маємо

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$f(x)$	3	9	10	13	5	15	11	16	14	8	7	4	12	2	6	1

Якщо задані число з інтервалу [1,16] і така таблиця, то відносно легко можна знайти прообраз x для $y = f(x)$. Але коли маємо число (наприклад 7) і такої таблиці немає, то знайти x для якого $f(x) = 7$ не є простою справою.

2) Ціле натуральне число n називається простим, якщо воно має лише два дільники – 1 і n.

Візьмемо прості числа $p = 48611$, $q = 53993$ і обчислимо $m = pq = 2624653723$. Нехай $X = \{1, 2, \ldots, m-1\}$. Визначимо функцію на множині X за допомогою рівності $f(x) = r_x$ для кожного $x \in X$, де r_x – остача від ділення x^3 на m. Наприклад, $f(2489991) = 1981394214$, оскільки $2489991^3 = 5881949859 \cdot m + 1981394214$. Обчислення значення $f(x)$ відносно просте, але обчислення значення x для $y = f(x) = 1981394214$ практично неможливе. ♠

Використання функцій з секретом дає можливість досить ефективно знаходити функцію обернену до даної бієктивної функції. Наприклад, в пункті 2) попереднього прикладу секретом є розклад числа $m = p \cdot q = 2624653723$. Тобто, числа p і q являють собою секрет.

Прикладом нестійкої NP-повної проблеми є окремий випадок **проблеми упаковки рюкзака**. Причин нестійкості є декілька.

Перша причина, як зазначалося вище, полягає в тому, що при встановленні NP-повноти розглядається найгірший випадок.

Друга причина полягає в складності встановлення нової, більш низької верхньої границі складності NP-проблеми, навіть якщо проблема є важкою для обчисленнь. Ця нова границя в кращому випадку залишається не доведеною.

Третя причина пов'язана з особливостями прикладної криптосистеми. Реальні криптосистеми являються результатом компромісів між користувачами такої системи.

2.4. Елементи теорії ймовірностей

Нехай S – деяка фіксована n-елементна множина. Над елементами із S виконуються експерименти такі, що при кожному з них можливі n несумісних і рівноможливих результатів $E_1, E_2, \ldots, E_n$. Слово "несумісний" означає, що коли в даному експерименті появляється результат E_i, то жоден з решти результатів в тому ж експерименті появитися не може. Кожний такий результат E_i називається **елементарною подією**. Крім елементарних подій розглядаються також випадкові події, які складаються з елементарних подій. Наприклад, елементарною подією є результат $E_1 = 3$ при підкиданні гральної кості, але нас може цікавити подія A появи числа не більшого 3 на грані кості. Подія A включає елементарні події $E_1 = 1, E_2 = 2, E_3 = 3$. Отже, $A = \{E_1, E_2, E_3\}$.

Означення 32. *Нехай $|S| = n$, $|A| = m$. Ймовірністю випадкової події A називається відношення кількості несумісних рівноможливих елементарних подій, які складають подію A (тобто, числа m), до кількості всіх можливих подій (тобто, до числа n). Ця ймовірність позначається $p(A)$ і дорівнює $\frac{m}{n}$. Множина S називається* **полем подій**.

З означення ймовірності випадкової події A випливає, що
1) $0 \le p(A) \le 1$, 2) $p(S) = 1$ і 3) $p(\emptyset) = 0$.

Подія $\emptyset$ називається **неможливою**, оскільки їй не відповідає жодна елементарна подія, а подія S називається **достовірною**. Якщо подія $p(A) = \frac{m}{n}$, то ймовірність того, що подія A не наступить дорівнює $\frac{n-m}{n} = 1 - p(A)$. Подія $\bar{A}$, яка означає, що подія A не наступить, називається *протилежною* до події A.

2.4.1. Властивості ймовірностей

Нехай $A \cup B$ означає появу однієї з подій A або B і яку називають *сумою подій*. А $A \cap B$ означає подію появи обох подій A і B одночасно, і яку називають *добутком подій*.

Правило додавання ймовірностей. Нехай A і B – події, тоді
1) $p(A \cup B) = p(A) + p(B) - p(A \cap B)$;
2) $p(A \cup B) = p(A) + p(B)$, якщо $A \cap B = \emptyset$, тобто випадкові події A і B несумісні;
3) якщо $\bigcup_{k=1}^{n} E_k = S$ і $E_i \cap E_j = \emptyset$ $(i \ne j)$, то $\sum_{k=1}^{n} p(E_k) = 1$.

Правило 1) обгрунтовується тим, що коли $p(A) = \frac{m_1}{n}$, $p(B) = \frac{m_2}{n}$ і $|A \cap B| = k$, то $p(A \cap B) = \frac{k}{n}$. Отже, $p(A \cup B) = \frac{m_1+m_2-k}{n} = = p(A) + p(B) - p(A \cap B)$.

Правило множення ймовірностей. Нехай подія A появляється m_1 разів серед n_1 рівноймовірних результатів першого експеримента, а подія B – m_2 разів серед n_2 рівноймовірних результатів другого експеримента. Тоді $p(A) = \frac{m_1}{n_1}$ і $p(B) = \frac{m_2}{n_2}$. Розглянемо тепер експеримент, який полягає в тому, що одночасно появляються обидві події. В такому випадку поле S складається з $n_1 n_2$ елементів, а події $A \cap B$ відповідатимуть $m_1 m_2$ результатів. Отже,

$$p(A \cap B) = \frac{m_1 m_2}{n_1 n_2} = \frac{m_1}{n_1} \cdot \frac{m_2}{n_2} = p(A)p(B).$$

Одержана рівність називається **правилом множення ймовірностей**. Це правило узагальнюється таким чином: нехай $A_1, A_2, \ldots, A_k$ – деякі взаємно незалежні події, тобто умови експеримента, з результатом якого пов'язана поява однієї з подій, ніяким чином не залежить від появи інших подій. В такому випадку

$$p(A_1 \cap A_2 \cap \ldots \cap A_k) = p(A_1)p(A_2) \cdots p(Ak).$$

Якщо події не є незалежними, то правило множення може не виконуватися. Нехай A – подія, яка означає, що з урни, в якій знаходиться m чорних і $n - m$ білих кульок, дістається чорна кулька, а подія B – дістається теж чорна кулька, після того, як з урни дістали одну кульку. Якщо перша кулька була чорною (відбулася подія A), то в урні залишилося $m - 1$ чорних кульок і $n - m$ білих. Тоді ймовірність $p(B) = \frac{m-1}{n-1}$. Якщо ж перша кулька була білою (відбулася подія $\bar{A}$), то $p(B) = \frac{m}{n-1}$. Як бачимо ймовірність події B залежить від того відбулася, чи ні подія A. Ймовірність події B, якщо ~~відомо, що~~ подія A відбулася, називається **умовною ймовірністю** події B за умови появи події A і позначається $p(B|A)$. Отже, в нашому прикладі $p(B|A) = \frac{m-1}{n-1}$.

Обчислення значення $p(B|A)$ виконується таким чином. Подія A може наступити в $N = m(n-1)$ випадках (першою була витягнута чорна кулька, а далі одна з $n-1$ кульок). Подія B наступає в $M = = m(m-1)$ випадках (обидві кульки були чорними). Тоді подія, яка нас цікавить, може наступити в $\frac{m(m-1)}{m(n-1)} = \frac{m-1}{n-1}$ випадках. Нехай кількість рівноймовірних випадків появи подій A і B дорівнює K, а кількість появи і події A і події B дорівнює M. Тоді ймовірність події $A \cap B$ – $p(A \cap B)$ – дорівнює $\frac{M}{K}$. Але $\frac{M}{K} = \frac{N}{K} \cdot \frac{M}{N}$, де $N = = m(n-1), M = m(m-1)$, і тоді $\frac{M}{N} = p(B|A)$ і $p(A) = \frac{N}{K}$. Отже, $p(A \cap B) = p(A) \cdot p(B|A)$, звідки дістаємо

$$p(B|A) = \frac{p(A \cap B)}{p(A)}.$$

Властивості умовної ймовірності випливають з означення цієї величини:

1) $0 \leq p(B|A) \leq 1$;

2) $p(B|A) = 1$, якщо $A \subseteq B$ і, зокрема, коли B достовірна подія;

3) $p(B|A) = 0$, якщо події A і B несумісні і, зокрема, коли B неможлива подія;

4) якщо $C \subset B$, то $p(B|A) \leq p(C|A)$;

5) якщо B і C несумісні події, то $p(C \cup B|A) = p(B|A) + p(C|A)$;

6) якщо $B_1, B_2, \ldots, B_k$ попарно несумісні події, то $p(B_1 \cup B_2 \cup \ldots \cup B_k|A) = p(B_1|A) + p(B_2|A) + \ldots + p(B_k|A)$;

7) $p(\bar{B}|A) = 1 - p(B|A)$.

Тепер можна дати строге означення незалежних подій.

Означення 33. *Події A і B називаються незалежними, якщо $p(B|A) = p(B)$.*

2.4.2. Випадкові величини

Якщо розглянути задачу про те, скільки викликів приходить від абонентів на телефонну станцію протягом певного проміжку часу, то число цих викликів не є сталою

величиною. Спостереження показують, що це число має значні коливання. Подібна ситуація спостерігається з числом викликів швидкої медичної допомоги і в багатьох інших задачах. Число значень, які може приймати випадкова величина, може бути скінченним, зліченним або незліченним. Значення можуть розподілятися дискретно або заповнювати інтервал щільно. В криптографії, як правило, застосовується дискретна область, в якій приймає свої значення випадкова величина. До дискретних областей відносять або скінченні, або нескінченні, але зліченні області. Для того, щоб задавати ймовірності значень випадкової величини, вводять поняття функції розподілу випадкової величини. Якщо ξ – випадкова величина і x – довільне дійсне число, то ймовірність того, що ξ набуде значення x, називається функцією розподілу ймовірностей випадкової величини ξ. Нехай O – деяка зліченна область.

Означення 34. *Функцією розподілу випадкової величини ξ на дискретній області O називається відображення $P : O \to \mathcal{D}$ таке, що $P[\xi = x_i] = p_i$, $(i = 1, 2, \ldots, |O|)$ і для якого виконуються наступні умови:*

а) $p_i \geq 0$; *б)* $\sum\limits_{i=1}^{|O|} p_i = 1$.

Таким чином, випадковою величиною називається величина, значення якої залежить від випадку і для якої визначена функція розподілу ймовірностей.

Розглянемо дві функції розподілу ймовірностей, які будуть зустрічатися далі.

Рівномірний розподіл має вигляд:

$$P[\xi = x_i] = \frac{1}{|O|}.$$

Приклад 2.4.1. а) Нехай $\Omega = \{a, b, c, \ldots, x, y, z\}$ – алфавіт англійської мови. Виберемо випадково довільний елемент x із A, дотримуючись рівномірного розподілу. Тоді ймовірність того, що вибраним елементом буде літера d дорівнює $\frac{1}{|A|} = \frac{1}{26}$.

б) Нехай O – множина невід'ємних чисел, які мають не більше k бітів у двійковому зображенні. Виберемо з множини O випадковий елемент, дотримуючись рівномірного розподілу. Покажемо, що ймовірність вибрати число, яке має k бітів, дорівнює $\frac{1}{2}$.

Розіб'ємо множину O на дві підмножини $O_1 = \{0, 1, 2, \ldots, 2^{k-1} - 1\}$ і $O_2 = \{2^{k-1}, 2^{k-1} + 1, \ldots, 2^k - 1\}$. Множина O_2 включає всі двійкові числа, які мають k бітів у своєму двійковому зображенні. Оскільки $|O_1| = |O_2| = \frac{|O|}{2}$, то

$$P[x \in O_2] = P[\bigcup_{i=2^{k-1}}^{2^k-1} [x = i]] = \sum_{i=2^{k-1}}^{2^k-1} P[x = i]] = \frac{|O_2|}{|O|} = \frac{1}{2}. \spadesuit$$

Біноміальний розподіл. Нехай ξ означає число появи події A в послідовності n незалежних дослідів, в кожному з яких ймовірність появи події A постійна і дорівнює p. В залежності від випадку ξ може приймати всі цілочисельні значення від 0 до n включно. Яка ймовірність того, що подія A появиться m разів в n дослідах, тобто чому дорівнює $P[\xi = m]$? Знайти цю ймовірність можна за допомогою таких підрахунків: в послідовності із n елементів нас цікавить m елементів, які означають появу події A. Всіх таких випадків буде очевидно C_n^m, де C_n^m – біноміальний коефіцієнт. Кожний випадок появи події A має ймовірність p, а кожний випадок не появи A (тобто появи події $\bar{A}$) має ймовірність $1 - p$. Отже,

$$P[\xi = m] = C_n^m p^m (1-p)^{n-m}.$$

Якщо випадкова величина ξ приймає значення з множини $O = \{0, 1, 2, \ldots, n\}$ і для величини $p \in (0, 1)$ виконується отримана вище залежність, то говорять, що випадкова величина ξ має **біноміальний розподіл ймовірностей**.

Приклад 2.4.2. Розглянемо такі задачі. Нехай під час виконання експеримента ідеальна монета підкидалася 10 разів. Яка ймовірність того, що

а) "орел" випаде п'ять разів?

б) "орел" випаде не більше п'яти разів?

Розв'язання. а) Застосовуючи функцію біноміального розподілу ймовірностей, знаходимо: $P[\xi = 5] = C_{10}^5 (\frac{1}{2})^5 (\frac{1}{2})^5 = C_{10}^5 (\frac{1}{2})^{10} = \frac{252}{1024} \approx 0,246$.

б) В цій задачі необхідно знайти суму ймовірностей всіх подій, за яких орел випаде не більше п'яти разів. Отже, маємо $P[\xi \leq 5] = (\frac{1}{2})^{10} \sum_{i=0}^{5} C_i^{10} = 0,623$. ♠

Ланцюги Маркова. Безпосереднім узагальненням схеми незалежних ~~досліджень~~ є схема ланцюгів Маркова. В цій схемі виконується послідовність експериментів, в кожному з яких може відбутися одна і тільки одна із k несумісних подій $A_1^s, A_2^s, \ldots, A_k^s$ (верхній індекс означає номер експеримента).

Означення 35. *Говорять, що послідовність експериментів утворює простий ланцюг Маркова, якщо умовна ймовірність в $(s+1)$-му експерименті $(s = 1, 2, \ldots)$ відбутися події A_i^{s+1} $(i = = 1, 2, \ldots, k)$ залежить тільки від того, яка подія відбулася в s-му експерименті і не змінюється від додаткових відомостей про те, які події відбувалися в раніше проведених есперимент ах.*

Далі обмежимося розглядом *однорідних ланцюгів Маркова*, в яких умовна ймовірність ~~відбутися~~ події A_j^{s+1} в $(s+1)$-му експерименті, не залежить від номера експеримента. Таку ймовірність називають ймовірністю переходу і позначають p_{ij}, де перший індекс означає результат попереднього експеримента, а другий індекс вказує, в який стан перейде система в наступний момент часу. Формальне означення має вигляд.

Означення 36. *Однорідний ланцюг Маркова - це послідовність випадкових величин $\{\xi_i\}, i = 1, 2 \ldots$, що приймають свої значення у дискретній множині Z така, що для будь-якого $n \geq 2$*

$$P(\xi_n = x_n | \xi_1 = x_1, \ldots, \xi_{n-1} = x_{n-1}) = P(\xi_n = x_n | \xi_{n-1} = x_{n-1}) = p(x_n x_{n-1}).$$

(Для неоднорідного марковського ланцюга ймовірності переходу $p(x_n x_{n-1})$ залежали б від n, а в даному випадку вони залежать лише від станів $x_{n-1} x_n$).

Повна ймовірнісна ситуація всіх можливих змін, які виконуються при переході від одного експеримента до безпосередньо наступного, визначається матрицею

$$\pi_1 = \begin{pmatrix} p_{11} & p_{12} & \ldots & p_{1k} \\ p_{21} & p_{22} & \ldots & p_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ p_{k1} & p_{k2} & \ldots & p_{kk} \end{pmatrix},$$

яка складається з ймовірностей переходів і яка називається *матрицею переходу*.

Властивості елементів матриці переходу зводяться до таких:

а) $0 \le p_{ij} \le 1$;

б) $\sum_{j=1}^{k} p_{ij} = 1$ $(j = 1, 2, \ldots, k)$.

А для самих матриць переходу виконується таке правило. Якщо

$$\pi_n = \begin{pmatrix} P_{11}(n) & P_{12}(n) & \ldots & P_{1k}(n) \\ P_{21}(n) & P_{22}(n) & \ldots & P_{2k}(n) \\ \ldots & \ldots & \ldots & \ldots \\ P_{k1}(n) & P_{k2})n) & \ldots & P_{kk}(n) \end{pmatrix},$$

матриця переходу після виконання n-го експеримента, то $\pi_n = \pi_1^n$.

Розглянемо тепер застосування ймовірносних методів до аналізу властивостей природних мов.

2.4.3. Ентропія і інформація

Основною властивістю випадкових подій є відсутність впевненості в тому, що вони відбудуться, а це створює певну невизначеність еквпериментів, пов'язаних з цими подіями. Зрозуміло, що степінь такої невизначеності в різних випадках буде різним. Але з практичної точки зору важливо вміти чисельно оцінювати степінь невизначеності результатів самих різноманітних експериментів, щоб мати можливість порівнювати їх між собою.

Нехай маємо k рівноймовірних результатів екпериментів. Очевидно, що степінь невизначеності кожного результату визначається числом k: при $k = 1$ результат експерименту не є випадковим, але коли k велике число, то передбачити результат важко. Отже, шукана числова характеристика степеня невизначеності є функцією $f(k)$ від k. Для детальнішого визначення функції $f(k)$ необхідно вимагати від неї виконання певних додаткових умов. Розглянемо два незалежних експерименти α і β. Нехай експеримент α має k рівноймовірних результатів, а експеримент β – l рівноймовірних результатів. Зрозуміло, що результат експеримента $\alpha\beta$ більш невизначений, ніж α і β тому, що до невизначеності експеримента α додається невизначеність експеримента β. Логічно вважати, що степінь невизначеності експеримента $\alpha\beta$ дорівнює сумі невизначеностей експериментів α і β, а оскільки експеримент може мати kl рівноймовірних результатів, то приходимо до такої умови, якій повинна задовольняти функція $f(k)$:

$$f(kl) = f(k) + f(l).$$

Функцією з такими властивостями є логарифм, так як $\log(kl) = = \log k + \log l$, а це означає, що за міру невизначеності експеримента, який має k рівноймовірних результатів приймається число $\log k$. Таке означення узгоджується з тим, що при $k = 1$ міра невизначеності $\log k = 0$ і що при зростанні k ця міра теж зростає.

Зазначимо, що вибір основи логарифмів несуттєвий, оскільки на підставі відомої формули $\log_b k = \log_a b \log_a k$ перехід від однієї основи до другої зводиться лише до множення функції $f(k) = \log k$ на константу (модуль переходу $\log_b a$). Тому далі буде використовуватися логарифм за основою 2. Це означає, що за одиницю виміру степеня

невизначеності експеримента, який має лише два рівноймовірних результати, приймається **біт**. Але це не обмежує загальності тому, що коли взяти за основу логарифму число 10 (десяткова одиниця або **дит**), то степінь невизначеності буде приблизно в $\log 10 \approx 3\frac{1}{3}$ рази більшим двійкової одиниці.

Нехай маємо експеримент, який може давати k рівноймовірних результатів $A_1, A_2, \ldots, A_k$ (тобто $p(A_i) = \frac{1}{k}$). Так як загальна невизначеність експеримента дорівнює $\log k$, то можна вважати, що кожний окремий результат, який має ймовірність $\frac{1}{k}$, вносить невизначеність $\frac{1}{k}\log k = -\frac{1}{k}\log\frac{1}{k}$. Аналогічно до цього можна покласти, що в загальному випадку, для експеримента з таблицею ймовірностей

експеримент	A_1	A_2	$\cdots$	A_k
ймовірність	$p(A_1)$	$p(A_2)$	$\cdots$	$p(A_k)$

міра невизначеності дорівнює

$$H(\alpha) = -p(A_1)\log p(A_1) - p(A_2)\log p(A_2) - \ldots - p(A_k)\log p(A_k).$$

Означення 37. *Число $H(\alpha)$ називається ентропією експеримента α.*

Властивості ентропії.

а) $H(\alpha) \geq 0$, оскільки $p(A_i) \geq 0$ і $-p(A_i)\log p(A_i) \geq 0$;

б) $H(\alpha) = 0$ тоді, коли $p(A_i) = 1$ $(i = 1, 2, \ldots, k)$;

в) якщо $p(A_i)$ мале число, то і $H(\alpha)$ мале число;

г) максимальна ентропія$H(\alpha)$ досягається при рівномірному розподілі ймовірностей, тобто

$$H(\alpha) = -\tfrac{1}{k}\log\tfrac{1}{k} - \tfrac{1}{k}\log\tfrac{1}{k} - \ldots - \tfrac{1}{k}\log\tfrac{1}{k} = \log k.$$

Умовна ентропія. У розглянутому вище випадку вважалося, що експерименти незалежні. Але в дійсності це не завжди так. Якщо експерименти залежні, то вже не можна бути певним того, що ентропія експеримента $\alpha\beta$ дорівнюватиме сумі $H(\alpha) + H(\beta)$.

Припустимо, що експерименти α і β не є незалежними. Розглянемо ентропію $H(\alpha\beta)$:

$$\begin{aligned} H(\alpha\beta) = {} & -p(A_1B_1)\log p(A_1B_1) - \ldots - p(A_1B_l)\log p(A_1B_l) - \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & -p(A_kB_1)\log p(A_kB_1) - \ldots - p(A_kB_l)\log p(A_kB_l), \qquad (2.1) \end{aligned}$$

де $A_1, A_2, \ldots A_k, B_1, B_2, \ldots, B_l$ означають можливі результати експериментів α і β. Тепер $p(A_1B_1) = p(A_1)p(B_1|A_1)$ де $p(B_1|A_1)$ – умовна ймовірність події B_1 за умови, що відбулася подія A_1. Отже,

$$\begin{aligned} & H(\alpha\beta) = -p(A_1)p(B_1|A_1)(\log p(A_1) + \log p(B_1|A_1)) - \\ & -p(A_1)p(B_2|A_1)(\log p(A_1) + \log p(B_2|A_1)) - \ldots - p(A_1)(p(B_l|A_1)(\log p(A_1) + \\ & + \log p(B_l|A_1)) = -p(A_1)(p(B_1|A_1) + p(B_2|A_1) + \ldots + p(B_l|A_1))\log p(A_1) + \\ & + p(A_1)(-p(B_1|A_1)\log p(B_1|A_1) - p(B_2|A_1)\log p(B_2|A_1) - \ldots - p(B_l|A_1)\log p(B_l|A_1)). \end{aligned}$$

Але $p(B_1|A_1) + p(B_2|A_1) + \ldots + p(B_l|A_1) = 1$ (оскільки подія B_1, $B_2, \ldots, B_l$ достовірна) і того, що

$-p(B_1|A_1)\log p(B_1|A_1) - \ldots - p(B_l|A_1)\log p(B_l|A_1) = H_{A_1}(\beta)$,

за умови, що відбулася подія A_1. Таким чином, перший рядок (2.1) приймає вигляд: $-p(A_1)\log p(A_1) + p(A_1)H_{A_1}(\beta)$. Аналогічно отримуємо вирази для решти рядків із (2.1). В результаті приходимо до виразу

$$H(\alpha\beta) = H(\alpha) + p(A_1)H_{A_1}(\beta) + p(A_2)H_{A_2}(\beta) + \ldots + p(A_k)H_{A_k}(\beta).$$

Покладаючи $H_\alpha(\beta) = p(A_1)H_{A_1}(\beta) + p(A_2)H_{A_2}(\beta) + \ldots + p(A_k)H_{A_k}(\beta)$, дістаємо

$$H(\alpha\beta) = H(\alpha) + H_\alpha(\beta).$$

Отриманий вираз наивається **правилом додавання ентропій**.

Властивості умовної ентропії. Умовна ентропія задовольняє такі властивості:

1) $H_\alpha(\beta) \geq 0$;

2) $H(\alpha\beta) = H(\alpha) + H_\alpha(\beta) = H(\beta) + H_\beta(\alpha)$.

Інтуїтивно це означає, що невизначеність експериментів α і β дорівнює невизначеності α плюс невизначеність β, після того, як результат експерименту α став відомим. Друга рівність в 2) має місце на підставі того, що α і β входять у формулу симетрично.

3) $H(\alpha) \geq H_\alpha(\beta) = H_\alpha(\beta\gamma)$.

Ця властивість означає, що додаткові знання не можуть збільшити невизначеність.

Елементи теорії інформації. Розглянемо спочатку моделі джерел відкритого тексту, які необхідні для математичного аналізу тексту та його криптографічних перетворень.

Під текстом будемо розуміти послідовність літер деякого скінченного алфавіту $X = \{x_1, x_2, \ldots, x_m\}$, де x_i – літери або символи цього алфавіту. Елементами алфавіту можуть бути власне літери; літери та цифри; літери, цифри та знаки пунктуації, взагалі скінченний набір будь-яких символів. Як правило, ми будемо розглядати український, російський чи латинський алфавіти (малі літери) зі знаком пропуску, що вважається літерою, або без нього, або ж двійковий алфавіт, що складається з двох символів: 0 та 1.

Відкритий текст (ВТ) – це текст, що підлягає шифруванню, шифрований текст (ШТ) – це текст, що утворюється в результаті шифрування. Відкритий та шифрований тексти можуть бути записані як у одному й тому ж, так і в різних алфавітах.

Нагадаємо, що словом довжини n (n-грамою) називається слово в алфавіті X, яке сладається з n символів цього алфавіту. При $n = 2$ це біграма, при $n = 3$ – триграма.

Будь-який текст має певну статистичну структуру. Для опису цієї структури користуються різноманітними ймовірносними моделями мови.

Нехай маємо деяке джерело, яке генерує відкритий текст. Це джерело відкритого тексту генерує послідовність символів алфавіту $x_1, x_2, \ldots, x_n, \ldots$ випадковим чином. Воно визначається алфавітом та ймовірностями появи n-грам: $P(\xi_{i+1} = x_1, \xi_{i+2} = x_2, \ldots, \xi_{i+n} = = x_n)$ для будь-яких цілих $n \geq 1, i \geq 0$ (тут $\xi_{i+1}, \xi_{i+2}, \ldots, \xi_{i+n}, \ldots$ – випадкові величини, а $x_1, x_2, \ldots, x_n$ – літери алфавіту X), які мають задовольняти умовам:

$$\sum_{x_1,\ldots,x_n\in X} P(\xi_{i+1}=x_1,\xi_{i+2}=x_2,\ldots,\xi_{i+n}=x_n)=1,$$

а також для будь-якого цілого $s\geq 1$

$$\sum_{x_1,\ldots,x_n\in X} P(\xi_{i+1}=x_1,\ldots,\xi_{i+n}=x_n,\xi_{i+n+1}=x_{n+1},\ldots,\xi_{i+n+s}=x_{n+s}) = P(\xi_{i+1}=x_1,\ldots,\xi_{i+n}=x_n).$$

Джерело називають стаціонарним, якщо для будь-яких цілих $n\geq 1, 1\leq i_1<\ldots<i_n, j\geq 0$ і будь-якого набору літер алфавіту $x_1,\ldots,x_n$ виконується рівність:

$$P(\xi_{i_1+j}=x_1,\ldots,\xi_{i_n+j}=x_n)=P(\xi_{i_1}=x_1,\ldots,\xi_{i_n}=x_n).$$

Далі будемо розглядати лише стаціонарні джерела, тобто такі, у яких немає залежності від зсуву j. Для стаціонарних джерел достатньо задати ймовірності $P(\xi_1=x_1,\xi_2=x_2,\ldots,\xi_n=x_n)$ для $n\geq 1$.

В залежності від властивостей сумісних розподілів $P(\xi_1=x_1,\xi_2=x_2,\ldots,\xi_n=x_n)$ для $n\geq 1$, можна побудувати різні моделі джерела ВТ.

Найбільш уживаними є описані нижче чотири моделі, з яких кожна наступна більш адекватно, ніж попередня, відображає структуру мови. Перша з них є простою й менш за всі враховує реальні статистичні властивості мови. Цю модель називають М0.

М0. У цій моделі джерело у кожен момент часу генерує символи із X незалежно та рівноймовірно:

$$P(\xi_i=x)=\tfrac{1}{m},\ i=1,2,\ldots,m, x\in X.$$

Всі n-грами в моделі М0 є рівноймовірними

$$P(\xi_i=x_1,\ldots,\xi_n=x_n)=\tfrac{1}{m^n}$$

для будь-яких цілих $n\geq 1$ та $x_1,x_2,\ldots,x_n\in X$.

Модель М0 має допоміжний характер, в зв'язку з тим, що вона не враховує навіть найпростіших властивостей мови. Наступна модель ВТ враховує частоти, з якими окремі літери алфавіту зустрічаються у мові.

М1. Символи тексту $x_1,x_2,\ldots,x_n\in X$ є незалежними, але вони генеруються з різними ймовірностями $P(\xi_i=x)=p(x), i=1,2,\ldots, x\in X$, де розподіл $p(x)$ відомий.

Розподіл ймовірностей $p(x)$ відповідає частотам появи літер $x\in\in X$ у мові. У цій моделі ймовірність появи n-грами має вигляд:

$$P(\xi_i=x_1,\ldots,\xi_n=x_n)=\prod_{i=1}^{n}p(x_i).$$

Зазначимо, що ймовірності появи літер у природних мовах значно відрізняються. Наприклад, найбільшу частоту в українській та російській мовах має літера "о", в англійській – літера "е". В російській мові літера "о" зустрічається майже в 50 разів частіше літери "ф", яка має найменшу частоту.

Наступна модель враховує мовну залежність між двома літерами, що стоять поряд.

М2. Джерело генерує біграми $x_1x_2, x_3x_4, x_5x_6\ldots$ незалежно одну від одної. Тобто на множині всіх біграм заданий розподіл ймовірностей $p(x_i,x_j)$ $i,j=1,\ldots,m$, і кожна

нова біграма джерела генерується незалежно від інших. У цій моделі ймовірність появи біграм має вигляд: $p(x_1x_2)p(x_3x_4)\dots p(x_{n-1}x_n)$.

Складніші залежності мови враховуються за допомогою марковської моделі.

М3. У цій моделі джерела ВТ послідовність $x_1, x_2, \dots$ утворює стаціонарний ланцюг Маркова. Для задання такого ланцюга достатньо задати розподіл початкових станів $p_0(x_i)$, $x_i \in \in X$ та перехідні ймовірності $p_{x_ix_j} = P(\xi_{n+1} = x_j|\xi_n = x_i), x_i, x_j \in \in X$, які на підставі однорідності не залежать від n.

При накладанні деяких умов на ланцюг Маркова (які не суперечать властивостям природних мов) існує граничний розподіл $\pi_{x_i} = \lim\limits_{n\to\infty} P(\xi_n = x_i|\xi_{n-1} = x_j)$, що не залежить від початкового стану $p_0(x_j)$. Він називається стаціонарним розподілом ймовірностей марковського ланцюга. У цій моделі ймовірність появи послідовності $x_1x_2\dots x_n$ має вигляд: $p(x_1)p(x_2|x_1)\dots p(x_n|x_{n-1})$. Граничні розподіли ймовірності задовольняють наступним рівностям

$$\begin{cases} \sum\limits_{x_i\in X} \pi_{x_i} = 1; \\ \sum\limits_{x_i\in X} \pi_{x_i} p(x_i, x_j) = \pi_{x_j}, \ x_j \in X. \end{cases}$$

Ймовірність n-грами у моделі М3 при стаціонарному режимі можна записати у вигляді

$$P(\xi_1 = x_1, \dots, \xi_n = x_n) = \pi_{x_1} p(x_1, x_2)p(x_2, x_3)\cdots p(x_{n-1}x_n).$$

Нехай $M = \{q_1, q_2, \dots, q_m\}$ – скінченна множина, на якій заданий розподіл ймовірностей $P = \{p(q_1), p(q_2), \dots, p(q_m)\}$. Пара (M, P) в теорії інформації називається **скінченним ансамблем**. Елементи $q_i \in M$ будемо називати **повідомленнями**. Часто говорять просто про ансамбль M, розуміючи під цим пару (M, P).

Інтуїтивно зрозуміло, що малоймовірне повідомлення несе в собі більше інформації, ніж більш ймовірне. К. Шеннон запропонував для виміру кількості інформації застосовувати функцію, яка відповідає цьому інтуїтивному уявленню, і яка зручна при обчисленнях.

Означення 38. *Власною інформацією повідомлення q_i називається величина $I(q_i) = -\log p(q_i) \geq 0$, а ентропією ансамблю (M, P) за всіма повідомленнями величина $H(M) = -\sum\limits_{i=1}^{m} p(q_i)\log p(q_i)$.*

Як зазначалося вище, величина $H(M)$ інтерпретується як невизначеність експерименту, в якому з ансаблю M вибирається одне повідомлення q_i, ймовірність якого $p(q_i), i = 1, 2, \dots, m$.

Розглянемо декартів добуток скінченних множин M та M_1, тобто множину пар (q, r), де $q \in M, r \in M_1$. Нехай на множині пар задано розподіл ймовірностей $p(q, r)$. Тоді говорять, що ансамблі M та M_1 задані сукупно. Розподіл $p(q, r)$ індукує розподіли на M та M_1:

$$p(q) = \sum_{r_j\in M_1} p(q, r_j) \text{ і } p(r) = \sum_{q_i\in M} p(q_i, r).$$

Ці розподіли дають можливість розглядати M і M_1 як окремі ансамблі.

Означення 39. *Сукупною ентропією ансамблів M і M_1 називається величина* $H(MM_1) = \sum\limits_{q,r} p(q,r)\log p(q,r)$.

Сукупно задані ансамблі M і M_1 називаються незалежними, якщо $\forall(q,r)(p(q,r) = p(q)p(r))$.

Звідси випливає, що коли M і M_1 – незалежні , то $H(MM_1) = H(M) + H(M_1)$.

Нехай відомий результат k-го експерименту M. Тоді умовна ентропія дорівнює:

$$H_M(r) = -\sum_{q\in M} p(r|q)\log p(r|q),$$

а умовна ентропія ансамблю M_1 відносно ансамбля M дорівнює:

$$H_M(M_1) = -\sum_{q\in M} p(q)\sum_{r\in M_1} p(r|q)\log p(r|q) = \\ = -\sum_{q\in M, r\in M_1} p(q,r)\log p(q,r).$$

Означення 40. *Взаємною інформацією ансамблів M та M_1 називається величина* $I(M;M_1) = H(M) - H_M(M_1)$. *При незалежних ансамблях M та M_1* $I(M;M_1)) = 0$.

Ентропія на символ джерела. Якщо X – алфавіт, то n-грама $(x_1, x_2, \ldots, x_n) \in X^n$ і n ансамблів задані сукупно розподілом n-грам $P(\xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_n = x_n)$ (джерело стаціонарне: немає залежності від розташування n-грами в тексті). Ентропія n-грами

$$H(X^n) = -\sum_{x_j\in X} p(x_1, x_2, \ldots x_n)\log p(x_1, x_2, \ldots, x_n). \tag{2.1}$$

Середня ентропія на один символ n-грами дорівнює $H_n = \frac{H(X^n)}{n}$. Для стаціонарних джерел існує границя H_∞ цієї величини:

$$\lim_{n\to\infty} \frac{H(X^n)}{n} = H_\infty$$

і ця границя називається **ентропією на символ джерела**.

Розглянемо, чому буде дорівнювати ентропія на символ джерела для різних моделей ВТ, які були введені раніше.

М0. $H(X^n) = nH(X) = n\log m$ (третя та четверта властивості ентропії).

М1. Внаслідок незалежності літер у тексті

$H_\infty = \lim\limits_{\infty} \frac{H(X^n)}{n} = \frac{nH(X)}{n} = H(X) = -\sum\limits_{x\in X} p(x)\log p(x)$.

М2. Розглядаються тексти довжиною $2n$:

$H_\infty = \lim\limits_{\infty} \frac{H(X^{2n})}{2n} = \frac{nH(X^2)}{2n} = \frac{H(X^2)}{2} = H_2$ (оскільки біграми незалежні).

М3. Можна довести, що для джерела, яке описується однорідним ланцюгом Маркова зі стаціонарним розподілом π_{x_i} та ймовірностями переходу $p(x_jx_j)$, $x_i, x_j \in X$,

$$H_\infty = -\sum_{x_i,x_j \in X} \pi_{x_i} \log p(x_i, x_j)$$

Величина H_n є n-им наближенням до H_∞. Зазначимо, що перші наближення H_1, H_2, H_3 ще дуже відрізняються від H_∞ (див. нижченаведену таблицю). В той же час обчислити H_n при великих значеннях n практично неможливо через величезну кількість можливих n-грам. Але можна розглянути умовну ентропію n-го символу тексту при умові, що відомі результати $n-1$ попередніх експериментів, тоді $H^n = H(x_n|x_1, x_2, \ldots, x_{n-1})$. В теорії інформації доводиться, що послідовність H^n має границю при $n \to \infty$ і ця границя збігається з границею послідовності H_n. Отже,

$$H_\infty = \lim_{n\to\infty} H(x_n|x_1 \ldots, x_{n-1}).$$

Ця рівність служить **другим означенням ентропії на символ джерела**.

	H_0	H_1	H_2	H_3	H_5	H_8
Англійська мова	4.76	4.03	3.32	3.10	2.1	1.9
Російська мова	5	4.35	3.52	3.01	–	–

Друге означення ентропії на символ джерела використовується для експериментальної оцінки H_∞ шляхом вгадування людиною наступної літери тексту.

Використовуючи друге означення H_∞, А. Н. Колмогоров експериментально оцінив для російськоє мови H^n при великих значеннях n. Виявилося, що після H^{30} значення H^n вже практично не змінюються, тобто дорівнюють H_∞, в той час як H_{15} ще істотно відрізняється від H_∞. Аналогічні результати були отримані і для інших європейських мов (роботи К. Шеннона та ін.).

Означення 41. *Надлишковістю на символ джерела називається величина* $R = 1 - \frac{H_\infty}{H_0}$, *де* $H_0 = \log |X| = \log m$.

H_0 дорівнює максимальній кількості інформації, яку може нести в собі один символ джерела, а H_∞ – кількість інформації, яку насправді несе в собі один символ. Таким чином, для достатньо великих n величина nR є середньою кількістю "зайвих" символів у тексті довжиною n, після втрати яких теоретично можна відновити текст. Надлишковість європейських мов знаходиться десь на рівні 60-80%. Але це не означає, що після випадкового видалення 60% символів (літер) завжди залишиться можливість відновити текст. Відкидати літера потрібно вибірково, використовуючи всі закономірності мови, і відновлювати також, використовуючи всі ці закономірності. Але практично врахувати всі закономірності неможливо. Експериментально встановлено: тільки до 25% літер можна видалити випадковим чином, щоб при цьому текст залишився придатним для відновлення.

Якби літери в тексті були незалежними та рівномірно розподіленими, то H_∞ дорівнювало б H_0, а надлишковість R була б нульовою. Але тоді будь-яка послідовність літер була б змістовним текстом, і навіть найменша помилка при передачі повідомлень призводила б до іншого змістовного тексту й не могла б бути визначена. При усному мовленні ми "ковтаємо" частину звуків, або виголошуємо їх нечітко, на письмі інколи робимо орфографічні помилки, та завдяки надлишковості мови все одно розуміємо один одного. Тож надлишковість – це природний механізм, що сприяє розумінню та протидії помилкам.

2.4.4. Частотна характеристика символів мови

Як зазначалося вище (модель **M1**), важливе місце в характеристиці природної мови займає частота появи літер в її словах Не всі літери алфавіту тієї, чи іншої природної мови появляються в словах з однаковою частотою. Ця обставина досить важлива і відіграє велику роль в криптоаналізі. Якщо текст зашифрований в даному алфавіті, то частота появи символів в зашифрованому тексті несе певну інформацію про ВТ. Маючи в розпорядженні частоту появи символів алфавіту даної мови, криптоаналітик іколи має можливість відтворити ВТ.

Нижче наводиться частота появи літер англійської мови.

Символ	Очікувана	Обчислена	Символ	Очікувана	Обчислена
A	7,9	7,5	N	7,0	7,0
B	1,5	1,4	O	7,4	7,5
C	3,0	4,1	P	3,0	3,0
D	4,0	3,2	Q	0,2	0,2
E	12,8	12,6	R	6,5	6,7
F	2,0	2,3	S	6,0	7,3
G	1,5	1,9	T	8,9	9,2
H	6,0	3,8	U	3,0	2,8
I	6,5	7,7	V	1,0	1,0
J	0,5	0,2	W	1,5	1,4
K	0,4	0,4	X	0,5	0,3
L	3,7	3,8	Y	2,0	1,6
L	0,0	0,0	Z	0,2	0,1
M	3,0	3,0			

Таблиця 1. Частота появи літер алфавіту англійської мови.

Приклад застосування частотного криптоаналізу. Нехай зашифрований текст в алфавіті англійської мови має вигляд [39]:

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXATZVUEPHZ
HMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSXEPYEPOPDZSZUFPOMBZ
WPFUPZHMDJUDTMOHMQ

На першому етапі можна обчислити відносну частоту появи літер алфавіту англійської мови на підставі таблиці 1. Якщо повідомлення має велику довжину, то застосування частотного аналізу є ефективним методом для криптоаналітика. Але оскільки маємо в розпорядженні текст не великої довжини, то не можемо надіятися на точне розв'язання задачі криптоаналізу. Відносна частота появи літер в даній шифрограмі має вигляд (в процентному відношенні):

Літера	Частота	Літера	Частота	Літера	Частота
P	13,13	E	5,00	B	1,67
Z	11,67	V	4,17	G	1,67
S	8,33	X	4,17	Y	1,67
U	8,33	F	3,33	I	0,83
O	7,50	W	3,33	J	0,83
M	6,67	Q	2,50	C	0,00
H	5,83	T	2,50	K	0,00
D	5,00	A	1,67	L,N,R	0,00

Таблиця 2. Відносна частота появи літер в шифрограмі

Порівнюючи ці дані з даними в таблиці 1 правдоподібним видається, що літери P і Z в шифрограмі відповідають літерам e і t в явному тексті, але у нас не має впевненості в тому, яка літера відпідає якій. Літери S, U, O, M і H виступають відносно часто і правдоподіним є їх відповідність літерам явного тексту із множини $\{r, n, i, o, a, s\}$.

Літери з найменшою частотою, тобто A,B,G,Y,I,L, правдоподібно належать до підмножини літер $\{w, v, b, k, x, q, j, z\}$.

В цьому місці маємо декілька можливих шляхів відтворення тексту явного.

Можемо продовжувати розпізнавати літери явного тексту, щоб вияснити, чи стає він деяким сенсовним текстом оригінального повідомлення. Біль систематичним методом є пошук чергових регулярностей. Наприклад, можливо певні слова знайдуться в тексті. Але можна шукати послідовності літер в шифрограмі, які повторюються, і пробувати і здогадатися про їх відповідників в явному тексті.

Ефективним способом аналізу є частота появи дволітерових комбінацій. Можна побудувати таблицю, подібну до таблиці 1, яка включає частоту появи в текстах дволітерових комбінацій. Найчастіше в англомовних текстах появляється комбінація th. В тексті нашої шифрограми найчастіше виступає комбінація ZW, яка появляється 3 рази. Тоді припускаємо, що літері Z відповідає літера *t*, а літері W – літера *h*. Прямуючи далі таким шляхом, припускаємо, що літері P відповідає літера e. В шифрограмі виступає комбінація літер ZWP, яку можемо перекласти як "the". Ця комбінація начастіше появляється в англомовних текстах, що показує, що ми прямуємо правильним шляхом.

Звернемо увагу на комбінацію літер ZWSZ в першому рядку. Не відомо, чи ці літери складають одне слово чи ні. Якщо це так, то воно має вигляд (наприклад that), а звідси випливає, що літері S відповідає літера a. Отже, маємо:

```
UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXATZVUEPHZ
 t a        e  e te  a that e e a       a   t   e t
HMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSXEPYEPOPDZSZUFPOMBZ
   ta t ha e ee  a e  th    t  a  e  e e tat  e   t
WPFUPZHMDJUDTMOHMQ
he  et
```

Зідентифікували ми тільки чотири літери, але маємо в розпорядженні не такий малий фрагмент відтвореного тексту. Продовження аналізу в такому напрямку шляхом проб і помилок приводить до повного відтворення явного тексту шифрограми. Повний відкритий текст, з доданими пропусками між словами, є таким:

```
it was disclosed yesterday that several informal but  direct
contacts have  been  made  with political representatives of
the viet cong in moscow.
```

♠

Контроль знань і вправи

1. Що таке джерело відкритого тексту? '
2. Чому розглядаються стаціонарні джерела і що це означає?
3. Які вам відомі моделі відкритого тексту?
4. Дайте визначення ентропії ансамблю.
5. Назвіть найважливіші властивості ентропії.
6. Які вам відомі означення ентропії на символ джерела?
7. Скільки доданків у правій частині формули (2.1)?
8. Що таке надлишковість на символ джерела? Чому приблизно дорівнює надлишковість європейських мов?
9. Навести означення кількості інформації, інформаційного вмісту мови та ентропії.
10. Що таке порядок росту функції? Які вам відомі оцінки порядку росту функцій? Дайте означення ін'єктивної, сюр'єктивної та бієктивної функцій.
11. Які вам відомі класи часової складності алгоритмів?
12. Яка функція називається односторонньою та односторонньою з секретом?
13. Упорядкувати нижченаведені функції за порядком зростання:
    а) 2^n, $\log_2 \log_2 n$, $n^2 + \log_2 n$, $\log_2 n$,
    б) $n - n^2 + 6n^3$, 2^{n-1}, n^2, n^3, $n \log_2 n$,
    в) $\sqrt{n}$, 8, n, $n!$, $(3/2)^n$, $(\log_2 n)^2$.
14. Для кожної пари нижченаведених функцій f і g виконується тільки одна з рівностей $f = O(g)$ або $g = O(f)$. Визначити, який із випадків має місце.
    а) $f(n) = (n^2 - n)/2$, $g(n) = 6n$, б) $f(n) = n + 2\sqrt{n}$, $g(n) = n^2$,
    в) $f(n) = n + n \log_2 n$, $g(n) = n\sqrt{n}$, г) $f(n) = n^2 + 3n + 4$, $g(n) = n^3$,
    д) $f(n) = n \log_2 n$, $g(n) = n\sqrt{n}/2$, е) $f(n) = 2 \log_2 n^2$, $g(n) = \log_2 n + 1$.
15. Навести приклади односторонніх функцій та функцій з класів NP і NPC.

2.5. Елементи загальної алгебри

Розглянемо основні поняття алгебраїчних структур, які використовуються в критографії.

2.5.1. Групи

Групи є однією із найуживаніших універсальних алгебр не тільки в комп'ютерних науках, але й в інших областях науки і техніки.

Нагадаємо означення універсальної алгебри, її підалгебри та ізоморфізму.

Означення 42. Універсальною Ω-алгеброю (*або просто* **алгеброю**) *називається система $G = (A, \Omega)$, яка складається з деякої непустої множини A* **(основна множина алгебри** *або* **носій алгебри)** *і множини визначених на A операцій $\Omega = \{\omega_1^{k_1}, \omega_2^{k_2}, \ldots, \omega_n^{k_n}, \ldots\}$ фіксованої арності* **(сигнатура алгебри)**, *де $k_i \in \in N$, $i = 1, \ldots, n, \ldots$. Операції із множини Ω називаються* **основними операціями алгебри.**

Нехай $G = (A, \Omega)$ – довільна алгебра, $\omega \in \Omega$ – n-арна операція і $A' \subseteq A$. Підмножина A' називається *замкнутою* відносно операції ω, якщо для довільних $a_1, \ldots, a_n$ із A' істинно $\omega(a_1, \ldots, a_n) \in A'$. Система (A', Ω) називається *підалгеброю алгебри* (A, Ω), якщо $A' \subseteq \subseteq A$ і A' замкнута відносно довільної основної операції алгебри $G = = (A, \Omega)$.

Означення 43. *Універсальні алгебри $G = (A, \Omega)$ і $Q = (B, \Omega')$ називаються алгебрами одного* **типу**, *якщо між елементами сигнатур Ω і Ω' можна встановити таку бієкцію, за якої довільна операція ω із Ω і відповідна їй операція ω' із Ω' мають одну і ту ж арність.*

Отже, можна вважати, що в алгебрах одного типу задана одна і та ж сигнатура операцій.

Означення 44. *Нехай алгебри $G = (A, \Omega)$ і $Q = (B, \Omega)$ одного типу. Якщо існує відображення $\varphi : A \to B$ таке, що для всіх елементів $a_1, \ldots, a_n$ із A і довільної n-арної операції ω із Ω має місце рівність*

$$\varphi(\omega(a_1, \ldots, a_n)) = \omega(\varphi(a_1), \ldots, \varphi(a_n)), \tag{2.2}$$

то відображення φ називається **гомоморфізмом**, *а алгебра G такою, що гомоморфно відображається в алгебру Q.*

Якщо гомоморфізм алгебри G в алгебру Q є ін'єкцією, то він називається **мономорфізмом**, *а якщо гомоморфізм алгебри G в алгебру Q є сюр'єкцією, то він називається* **епіморфізмом**.

Якщо гомомрфізм φ – бієкція алгебри G на алгебру Q, то він називається **ізоморфізмом**, *а алгебри G і Q –* **ізоморфними** *($G \sim Q$).*

Гомоморфізм алгебри G в алгебру G (тобто $\varphi : G \to G$) називається **ендоморфізмом**, *а коли φ ізоморфізм G на G, то він називається* **автоморфізмом**.

Означення 45. *Універсальна алгебра $G(A,\Omega)$ називається групою, якщо множина Ω включає бінарну операцію множення $(\cdot)$, унарну операцію взяття оберненого елемента $(^{-1})$ і нульарну операцію (e), яка фіксує одиницю групи, і операція множення задовольняє законам асоціативності $((x\cdot y)\cdot z = x\cdot(y\cdot z))$, скорочення $(x\cdot x^{-1} = x^{-1}\cdot x = e)$ і законам для одиниці $(x\cdot e = e\cdot x = x)$, де $x,y,z\in A$.*

Символи x і x^{-1} називаються *взаємно оберненими*. Елементи групи утворюють вирази, які часто називають словами. Таке слово називається *нескорочуваним*, якщо воно не має жодної пари взаємно обернених символів, які стоять поруч. Довільне слово можна перетворити до нескорочуваного слова в результаті послідовного викреслювання пар взаємно обернених символів, які стоять поруч. Очевидно, що одиниця групи e буде нескорочуваним словом. Таке перетворення слів називається їх *приведенням*.

Тотожні співвідношення групи дозволяють записати довільне слово групи у вигляді нескорочуваного слова

$$x_{i_1}^{n_1}x_{i_2}^{n_2}\dots x_{i_k}^{n_k}, \tag{2.3}$$

де n_j – ціле число, x_{ij} не обов'язково всі різні, $j = 1,2,\dots,k$, а x^n - слово довжини n, яке визначається таким чином:

$$x^n = \begin{cases} \underbrace{xx\dots x}_{n \text{ разів}}, & \text{якщо } n>0;\\ e, & \text{якщо } n=0;\\ \underbrace{x^{-1}x^{-1}\dots x^{-1}}_{n \text{ разів}}, & \text{якщо } n<0. \end{cases}$$

Безпосередньо з аксіом групи випливають такі прості наслідки:

а) одиничний елемент в групі єдиний;

б) обернений елемент до даного елемента єдиний;

в) рівняння $ax = b$ в групі має єдиний розв'язок.

У випадку а) припустимо, що в групі існує принаймні два одиничні елементи e і e'. Тоді, на підставі їх властивості дістаємо $e = e\cdot e' = e'$. Отже, $e = e'$.

У випадку б) з означення групи випливає існування таких елементів a' і a", що $a\cdot a' = e$ і $a"\cdot a = e$. Але тоді на підставі закону асоціативності дістаємо: $a"aa' = a"(aa') = a"\cdot e = a"$ і $a"aa' = (a"a)a' = e\cdot a' = a'$. Звідки отримуємо, що $a' = a"$.

У випадку в) нехай c – розв'язок рівняння $ax = b$, тоді дістаємо тотожність $ac = b$. Домножаючи обидві частини цієї тотожності зліва на a^{-1} дістаємо: $c = a^{-1}(ac) = a^{-1}b$. Якщо d – деякий інший розв'язок рівняння $ax = b$, то виконуючи ті самі дії, що і вище, дістаємо $d = a^{-1}b$. А це означає, що $c = d$.

Приклад 3.1.1. Розглянемо множину A = {а,б,в,г,д,е} з операцією множення, яка задана такою таблицею:

*	а	б	в	г	д	е
а	е	д	г	в	б	а
б	в	г	д	е	а	б
в	б	а	е	д	г	в
г	д	е	а	б	в	г
д	г	в	б	а	е	д
е	а	б	в	г	д	е

.

Покажемо, що дана алгебра є групою, яку будемо позначати $G_з$.

З таблиці множення випливає, що роль одиниці відіграє елемент 'е'. Крім того, оскільки таблиця множення не симетрична відносно головної діагоналі, то операція множення не є комутативною.

Оберненими елементами для елементів 'а,б,в,г,д' є елементи 'а,г,в,б,д' відповідно. Отже, елементи 'а,в,д' обернені самі до себе, а елементи 'б' і 'г' – взаємно обернені.

Залишається показати, що операція множення задовольняє закон асоціативності, тобто $\forall x, y, z \in A$ $x(yz) = (xy)z$. Це зводиться до розгляду окремих випадків, яких за основним правилом комбінаторики буде $6^3 = 216$. Але для одиничного елемента немає потреби розглядати випадки, оскільки вони очевидним чином випливають безпосередньо з таблиці множення. Отже, кількість випадків зменшується до $5^3 = 125$.

Розглянемо випадки виразів, які включають елементи, обернені самі до себе: а(аx) = (аа)x = еx = x. Оскільки аx = е, д, г, в, б, а, якщо x = а, б, в, г, д, е відповідно, то рівність лівої і правої частин має місце. Аналогічно перевіряється і для 'в, д'.

Решту випадків та доведення того, що оберненим елементом до елемента xy буде елемент $y^{-1}x^{-1}$, пропонуються читачеві. ♠

Нехай H – деяка підгрупа групи G і a – довільний елемент із G. Множина $aH = \{ah | h \in H\}$ називається *лівим суміжним класом групи G по підгрупі H*, заданим елементом a. Ясно, що $a \in aH$, оскільки $e \in H$, і якщо $b \in aH$, то $bH = aH$. Дійсно, $b \in aH$ означає що $b = ah$, де $h \in H$. Але тоді для довільних $h_1, h_2 \in H$ маємо $bh_1 = (ah)h_1 = a(hh_1) \in aH$, тобто $bH \subseteq aH$. Нехай тепер $b = ah \in aH$, тоді для довільного $h_2 \in H$ дістаємо

$$ah_2 = (bh^{-1})h_2 = b(h^{-1}h_2) \in bH,$$

тобто $aH \subseteq bH$. Отже, $aH = bH$.

Звідси випливає, що ліві суміжні класи $H, a_1H, a_2H, \ldots, a_nH, \ldots$ є класами розбиття групи G. Отже, підгрупа H задає на G відношення еквівалентності R : $aRb \Leftrightarrow aH = bH$.

Задання групи G у вигдяді об'єднання класів $H \cup a_1H \cup a_2H \cup \cup a_3H \cup \ldots \cup a_nH \cup \ldots$ називається *лівостороннім розкладом групи G* за підгрупою H. Аналогічно будується і правосторонній розклад групи G за підгрупою H. Правосторонній і лівосторонній розклади групи складаються із одного і того ж числа класів. У цьому легко переконатися, задаючи відображення $f : G \to G$ так, що $f(a) = a^{-1}$.

Якщо ж число суміжних класів в розкладі групи G за підгрупою H скінченне, то підгрупа Н називається підгрупою **скінченного індексу** а число класів – **індексом** підгрупи H в групі G.

Підгрупа H називається **нормальним дільником**, або **інваріантною підгрупою** групи G, якщо лівосторонній розклад G по H збігається з правостороніим розкладом G за H. Інакше кажучи, для довільного $a \in G$ має місце рівність $aH = Ha$ або $aHa^{-1} = H$. Очевидно, що нормальними дільниками довільної групи буде одинична підгрупа, яка складається лише з одиниці групи, і сама група. Ці нормальні дільники називаються *тривіальними*.

Розклад групи G за підгрупою $H \subseteq G$, як було показано вище, задає відношення еквівалентності на G. Оскільки розклад може бути лівостороннім і правостороннім, то цих еквівалентностей буде дві – R_1 і R_2. Означення цих еквівалентностей має вигляд: $\forall a, b \in G$

$$(a, b) \in R_1 \Leftrightarrow a^{-1}b \in H \text{ і } (a, b) \in R_2 \Leftrightarrow ab^{-1} \in H.$$

Еквівалентності R_1 і R_2, як неважко переконатися, є конгруентностями і якщо підгрупа H – нормальний дільник групи G, то обидві конгруенції збігаються, тобто $R_1 = R_2$. Має місце і обернене твердження: якщо R відношення конгруентності на G, то клас розбиття, якому належить одиниця групи, є її нормальним дільником. В загальному випадку має місце така

Теорема 8. *В довільній групі G її конгруентності знаходяться у взаємно однозначній відповідності з її нормальними дільниками.*

Приклад 3.1.2. Повернемося до групи G_s з прикладу 3.1.1. Безпосередньо з таблиці множення цієї групи можна помітити, що вона має підгрупу H, носій якої складається з елементів 'е,б,г'. Дійсно, ця множина елементів замкнута відносно операцій групи:

*	е	б	г
е	е	б	г
б	б	г	е
г	г	е	б

З симетричності таблиці множення випливає закон комутативності для цієї підгрупи. Лівосторонній розклад групи G_s за підгрупою H має вигляд:

$$G_s = H \cup \text{а}H = \{\text{е,б,г,}\} \cup \{\text{а, д, в}\} = \{\text{е,а,б,в,г,д}\},$$

оскільки елементи д, в є елементами аH (в чому легко переконатися аH = {а, д, в} = = дH = {в, а, д} = вH = {д, в, а}).

Неважко переконатися, що правосторонній розклад збігатиметься з лівостороннім розкладом. Звідси випливає, що підгрупа H – нормальний дільник групи G_s, оскільки аHа = вHв = дHд = H. ♠

Нехай G і G' – групи, e і e' – їх одиничні елементи відповідно, а h – гомоморфізм групи G в групу G'. Тоді множина $K = \{a \in G | h(a) = e'\}$ називається **ядром гомоморфізму** h і позначається символом $ker(h)$. Неважко довести, що $ker(h)$ – нормальний дільник групи G (див. вправу 15).

Якщо група скінченна, то число її елементів називається *порядком групи* і, очевидно, всі її підгрупи теж будуть скінченними. Нехай G – скінченна група n-го порядку і H – її підгрупа k-го порядку. Тоді в розкладі G по H довільний суміжний клас складається в точності із k елементів і, отже, $n = kj$. Звідси випливає

Теорема 9 (Лагранж). *Порядок і індекс довільної підгрупи скінченної групи є дільниками порядка групи.*

Якщо G – група і $a \in G$, то покладемо $a^0 = e$ і $a^{-n} = (a^{-1})^n$, а додатні степені елемента a як $a^n = a \cdot a \cdot \ldots \cdot a$, де кількість множників дорівнює n.

Сукупність елементів виду a^n, де $n \in \mathcal{Z}$, називається *циклічною підгрупою* групи G. Степені елемента a не обов'язково повинні бути різними. Елемент a групи G називається **елементом скінченного порядку**, якщо існують такі цілі числа k і l, що $a^k = a^l$, тобто $a^{k-l} = e$. Найменший показник серед всіх таких показників елемента a називається *порядком елемента* a. Якщо таких чисел k і l не існує, то a має нескінченний порядок.

Приклади циклічних груп. а) Для $n \geq 1$ адитивна група лишків Z_n за модулем n є циклічною групою. Породжуючим елементом цієї групи, очевидно, є одиниця.

б) Мультиплікативна група Z_5 породжується елементами 2 і 3, оскільки

$$2^1 = 2,\ 2^2 = 4,\ 2^3 = 3,\ 2^4 = 1,$$
$$3^1 = 3,\ 3^2 = 4,\ 3^3 = 2,\ 3^4 = 1.$$

Елемент 4 не може бути породжуючим, оскільки $4^x \equiv 2\ (mod\ 5)$ не має розв'язку в цій групі.

в) Адитивна група Z_{12}, яка складається з чисел циферблату годинника. Породжуючими елементами цієї групи будуть числа 1, 5, 7, 11. Наприклад,

$$5 \cdot 5 = 1,\ 5 \cdot 10 = 2,\ 5 \cdot 3 = 3,\ 5 \cdot 8 = 4,\ 5 \cdot 6 = 6,$$
$$5 \cdot 11 = 7,\ 5 \cdot 4 = 8,\ 5 \cdot 9 = 9,\ 5 \cdot 2 = 10,\ 5 \cdot 7 = 11.$$

Порядок елемента 5 дорівнює 5, оскільки $5 \cdot 5 = 1$. З тих же причин порядок елемента 7 дорівнює 7, а порядок елемента 8 нескінченний. ♠

Якщо a – елемент скінченного порядку, то $a^0 = e, a, a^2, \ldots\ldots, a^{n-1}$ будуть різними елементами групи. Якщо дано a^k, де $k > n$, то $k = qn + r$, $0 \leq r < n$, і $a^{qn+r} = (a^n)^q \cdot a^r = a^r$. Отже, скінченний порядок елемента збігається з порядком його циклічної групи.

Наслідок 1. *Порядок довільного елемента скінченної групи є дільником порядка групи.*

Наслідок 2. *Довільна скінченна група, порядок якої є простим числом, буде циклічною.*

Дійсно, на підставі простоти порядку групи вона повинна збігатися з циклічною підгрупою, породженою довільним її елементом відмінним від одиниці. ■

Група називається *простою*, якщо вона не має нетривіальних нормальних дільників.

Якщо група H – нормальний дільник групи G, то множина суміжних класів складає групу. Дійсно, якщо $H, a_1H, a_2H, \ldots$ – розбиття групи G і для довільного елемента a із G мають місце рівності $aH = Ha$ і $H \cdot H = H$, то неважко довести, що:

1) $(aH \cdot bH) \cdot cH = aH \cdot (bH \cdot cH)$;

2) $aH \cdot H = H \cdot aH = aH = Ha$, тобто H відіграє роль одиниці;

3) $aH \cdot (a^{-1})H = (aHa^{-1}) \cdot H = H \cdot H = H$;

4) $aH \cdot bH = (Ha) \cdot (bH) = (H(ab)) \cdot H = abH \cdot H = Hab$.

Іншими словами, довільний нормальний дільник H групи G задає деяке відношення конгруентності R на G. Фактор-група за цим відношенням конгруентності позначається G/H і називається **фактор-групою групи** G за нормальним дільником H.

Якщо група G – скінченна і H – її нормальний дільник, то одержуємо

Наслідок 3. *Порядок групи G/H дорівнює індексу підгрупи H в групі G і є дільником порядка групи G.*

2.5.2. Групи підстановок

Розглянемо приклад некомутативної групи, яка відіграє важливу роль в теорії груп, – групу підстановок деякої скінченної множини $M = \{1, 2, , \ldots, n\}$. Підстановкою називається бієктивне відображення множини M на себе.

З попереднього відомо, що добуток відображень є операцією асоціативною.

Сукупність всіх підстановок множини M складає групу. Дійсно, роль одиниці в цій групі відіграє тотожна підстановка, яка залишає на місці кожний елемент із M. Далі, якщо елемент a переходить в елемент $f(a)$, то $f^{-1}(f(a)) = a$ і f^{-1} теж буде підстановкою на підставі взаємної однозначності f. f^{-1} буде відігравати роль оберненого елемента для f. Отже, всі підстановки множини M складають групу, яка називається **симетричною групою на множині** M.

Приклад 3.1.3. Якщо множина M скінченна і складається з n елементів, то симетрична група на M, яка називається **симетричною групою n-го степеня**, буде скінченною і матиме порядок $n!$ (див. вправу 14 в кінці розділу). Підстановки скінченних множин задаються у вигляді таблиць відповідностей. Наприклад, якщо $M = \{1,2,3,4,5,6\}$, то підстановки f і f_1 множини M, відповідні значення яких $\{4,3,1,5,2,6\}$ і $\{6,3,5,2,4,1\}$, матимуть вигляд

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 1 & 5 & 2 & 6 \end{pmatrix} \quad f_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 3 & 5 & 2 & 4 & 1 \end{pmatrix}.$$

Множення підстановок $f \cdot f_1$ є суперпозицією цих відображень і тоді

$$f \cdot f_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 1 & 5 & 2 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 3 & 5 & 2 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 5 & 6 & 4 & 3 & 1 \end{pmatrix}.$$

Нехай маємо підстановки f_1, f_2, f_3, які задані наведеними нижче таблицями. Для прикладу покажемо, що $(f_1 \cdot f_2) \cdot f_3 = f_1 \cdot (f_2 \cdot f_3)$:

$$f_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 5 & 2 & 6 & 1 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 2 & 6 & 1 & 4 \end{pmatrix},$$

$$f_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 3 & 5 & 1 & 4 & 2 \end{pmatrix}.$$

Тоді

$$(f_1 \cdot f_2) \cdot f_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 6 & 4 & 1 & 5 \end{pmatrix}, \quad f_1 \cdot (f_2 \cdot f_3) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 6 & 4 & 1 & 5 \end{pmatrix}.$$

Розглянемо приклад задання групи G' всіх підстановок множини $M = \{1,2,3\}$ за допомогою таблиці Келі. У цьому випадку група G' – скінченна алгебра 6-го порядку і її елементами є підстановки

$$f_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \quad f_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix},$$

$$f_4 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \quad f_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad f_6 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}.$$

Таблиця множення цієї алгебри набуває вигляду

$\cdot$	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	3	1	6	4	5
3	3	1	2	5	6	4
4	4	5	6	1	2	3
5	5	6	4	3	1	2
6	6	4	5	2	3	1,

де $h(f_i) = i,\ 1 = 1,2,3,4,5,6$.

Таблиця операції взяття оберненого елемента вже знаходиться в таблиці множення, тому наводити її не має сенсу.

Пропонується переконатися самостійно, що отримана група ізоморфна групі G з прикладу 3.1.1, якщо відображення $\varphi : G \to G'$ є таким:

$$\varphi(\text{а}) = 4,\ \varphi(\text{б}) = 3,\ \varphi(\text{в}) = 5,\ \varphi(\text{г}) = 2,\ \varphi(\text{д}) = 6, \varphi(\text{е}) = 1. \spadesuit$$

Окремим випадком підстановки є **інволюція**.

Означення 46. *Підстановка p скінченної множини S називається інволюцією, якщо $p = p^{-1}$ або $p(p(x)) = x$ для довільного $x \in S$.*

Приклад інволюції. Підстановка $p : S \to S$, де $S = \{1, 2, 3, 4, 5\}$,

$$p = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 5 & 1 & 3 \end{pmatrix}$$

є інволюцією. $\spadesuit$

Важливість груп підстановок випливає з такої теореми.

Теорема 10 (Келі). *Для довільної групи $G = (A, \Omega, E)$ існує ізоморфна їй група підстановок $G' = (M, \Omega, E)$ на деякій підходящій множині M.*

Нехай $f : M \to M$ – підставновка n-елементної множини M. Кожна така підстановка індукує відношення еквівалентності R на M:

$$xRy \Leftrightarrow (\exists k \in \mathcal{Z})y = f^k(x).$$

Дійсно воно рефлексивне, оскільки $x = f^0(x) = \varepsilon(x)$, де ε – тотожна підстановка.

Симетричність випливає з того, що коли $y = f^k(x)$, тобто xRy, то $x = f^{-k}(y)$. А це означає, що yRx.

Транзитивність R випливає з того, що коли xRy і yRz, тобто $y = f^k(x)$ і $z = f^m(y)$, то $z = f^{k+m}(x)$. А це означає, що xRz.

Відношення R розбиває множину M на класи еквівалентності:

$$M = M_1 \cup M_2 \cup \ldots \cup M_k, \tag{2.4}$$

де $M_i \cap M_j = \emptyset$ коли $i \neq j$, $i, j = 1, 2, \ldots, k$.

Означення 47. *Класи еквівалентності M_i за відношенням R називаються* **орбітами** *підстановки f.*

Очевидно, $f(M_1) = M_1, \ldots, f(M_k) = M_k$ і звуження $f_1, \ldots, f_k$ підстановки f на орбіти $M_1, \ldots, M_k$ теж будуть підстановками.

Означення 48. *Звуження f_i підстановки f на множину M_i називається* **циклом** *цієї підстановки.*

Цикл є підстановкою степеня $m_i = |M_i|$ на множині M_i, тобто $f^{m_i}(x) = x$, $x \in M_i$, $i = 1, 2 \ldots, k$, і його можна записати у вигляді

$$(x, f_i(x), f_i^2(x), \ldots, f_i^{m_i - 1}(x)).$$

Це дає змогу розкласти довільну підстановку f на цикли

$$(x_1, f_1(x_1), \ldots, f_1^{m_1-1}(x_1)) \ldots (x_k, f_k(x_k), \ldots, f_k^{m_k-1}(x_k)).$$

Розширимо підстановки $f_1, f_2, \ldots, f_k$ до підстановок степеня n, покладаючи $f_i(x) = x$, якщо $x \in M \setminus M_i$. Тоді розклад підстановки f можна подати у вигляді добутку циклів

$$f = f_1 f_2 \ldots f_k.$$

Цей запис однозначний з точністю до циклічних перестановок елементів у циклах і не залежить від порядку співмножників.

Означення 49. *Підстановка f на множині M має порядок p, якщо $f^p = \varepsilon$ і це число p найменше серед всіх таких чисел.*

Теорема 11. *Порядок p підстановки f дорівнює найменшому спільному кратному довжин циклів у розкладі f.*

Доведення. Якщо розклад підстановки f має вигляд $f = f_1 f_2 \ldots f_k$, то

$$f^p = f_1^p f_2^p \ldots f_k^p.$$

Оскільки $f_i^{m_i}(x) = x$ $(i = 1, 2, \ldots, k)$, то $p = \text{НСК}(m_1, m_2, \ldots, m_k)$. ■

Означення 50. *Підстановка f на множині M належить цикловому класу $\{1^{\alpha_1} 2^{\alpha_2} \ldots n^{\alpha_n}\}$, якщо вона має α_1 циклів довжини 1, α_2 циклів довжини 2,..., α_n циклів довжини n $(1 \cdot \alpha_1 + 2 \cdot \alpha_2 + \ldots + n \cdot \alpha_n = n)$.*

Нехай $P(\alpha_1, \alpha_2, \ldots, \alpha_n)$ – число підстановок в цикловому класі $\{1^{\alpha_1} 2^{\alpha_2} \ldots n^{\alpha_n}\}$.

Теорема 12.

$$P(\alpha_1, \alpha_2, \ldots, \alpha_n) = \frac{n!}{1^{\alpha_1} 2^{\alpha_2} \ldots n^{\alpha_n} \alpha_1! \alpha_2! \ldots \alpha_n!}. \tag{2.5}$$

Доведення. Розглянемо розклад підстановки f з циклічного класу $\{1^{\alpha_1} 2^{\alpha_2} \ldots n^{\alpha_n}\}$ у вигляді добутку циклів

$$f = (x_1)(x_2) \ldots (x_{\alpha_1})(y_1, z_1)(y_2, z_2) \ldots (y_{\alpha_2}, z_{\alpha_2}) \ldots.$$

Довільну підстановку з цього класу можна одержати шляхом всіх можливих перестановок елементів при збереженні дужок. Циклічні перестановки елементів всередині дужок і перестановки, які переводять повністю елементи з однієї дужки в іншу з такою самою кількістю елементів, очевидно, не генерують нових підстановок. Число таких підстановок дорівнює $\{1^{\alpha_1} 2^{\alpha_2} \ldots n^{\alpha_n}\}$. Помноживши цю кількість на число різних підстановок у цикловому класі $\{1^{\alpha_1} 2^{\alpha_2} \ldots n^{\alpha_n}\}$, отримуємо кількість всіх підстановок $n!$. Отже,

$$P(\alpha_1, \alpha_2, \ldots, \alpha_n) 1^{\alpha_1} 2^{\alpha_2} \ldots n^{\alpha_n} \alpha_1! \alpha_2! \ldots \alpha_n! = n!.$$

Звідси отримуємо

$$P(\alpha_1, \alpha_2, \ldots, \alpha_n) = \frac{n!}{1^{\alpha_1} 2^{\alpha_2} \ldots n^{\alpha_n} \alpha_1! \alpha_2! \ldots \alpha_n!}. \blacksquare$$

Нехай $C(n,k)$ означає число підстановок степеня n n-елементної множини, які мають k циклів. З попередньої теореми випливає, що

$$C(n,k) = \sum_{1\alpha_1+\ldots+n\alpha_n=n,\ \alpha_1+\ldots+\alpha_n=k,\ \alpha_i\geq 0} = \frac{n!}{1^{\alpha_1} 2^{\alpha_2} \ldots n^{\alpha_n} \alpha_1! \alpha_2! \ldots \alpha_n!}.$$

Можна показати, що $|C(n,k)| = S_n^k$, де S_n^k числа Стірлінга першого роду.

Застосування груп підстановок в криптографії. Нехай $X = \{a, b, c, \ldots, x, y, z\}$ – алфавіт англійської мови, літери якого лінійно упорядковані звичайним чином:

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Розглянемо приклади шифрування, які ґрунтуються на техніці підстановок.

а) **Шифр Цезаря**. Цей шифр відомий з давніх часів і приписують його авторство Юлію Цезарю – римському імператору. Цей шифр базується на заміні кожної літери алфавіту новою літерою цього ж алфавіту, яка знаходиться на три позиції правіше.

a	b	c	d	e	f	g	h	i	j	k	l	m
D	E	F	G	H	I	J	K	L	M	N	O	P
n	o	p	q	r	s	t	u	v	w	x	y	z
Q	R	S	T	U	V	W	X	Y	Z	A	B	C

Дано *відкритий текст:* "meeting will in twelve"
криптограма: PHHWLQJ ZLOO LQ WZHOYH
Зауважимо, що алфавіт "закручений" так, що після літери Z наступною буде йти літера A.

Якщо кожній літері припишемо числовий відповідник ($n(a) = 0, n(b) = 1, \ldots$), то цей шифр можна подати у такому вигляді. Кожна літера відкритого тексту p замінюється літерою тексту зашифрованого q на підставі правила:

$$q = f(p) = n(p) + 3 \ (mod\ 26).$$

Зсув літер в алфавіті може мати довільну величину, а це означає, що загальний вигляд алгоритма є таким:

$$q = f(p) = n(p) + k \ (mod\ 26),$$

де $k \in \{1, 2, \ldots, 26\}$. Алгоритм дешифрації досить простий:

$$p = g(q) = n(q) - k \ (mod\ 26).$$

Якщо відомо, що даний текст зашифрований шифром Цезаря, то його криптоаналіз не складає труднощів. Цей криптоаналіз можна виконати як методом частотного аналізу, так і методом простого перебору, випробовуючи 25 можливих ключів.

Нижче на рисунку показані результати застосування методу перебору до зашифрованого тексту. В даному випадку відкритий текст дістаємо на третьому кроці перебору.

Ключ	PHHWLQJ	ZLOO	LQ	WZHOYH
1	oggvkpi	yknn	kp	vygnxg
2	nffujog	xjmm	jo	uxfmwf
3	meeting	will	in	twelve
4	lddshmc	vhkk	hm	svdkud
⋮	⋮	⋮	⋮	⋮
25	...	...	...	...

Рис. 3.2.1. Криптоаналіз шифру Цезаря методом перебору

Застосування методу перебору стало можливим, оскільки:

1) відомий алгоритм шифрування і дешифрування;

2) існує тільки 25 можливих ключів;

3) мова відкритого тексту відома і легко розпізнається.

Метод підстановки стає практичним, якщо існує великий простір для вибору ключів. Наприклад, американський стандартний алгоритм DES використовує 56-бітовий ключ, що дає простір вибору 2^{56}, або більше $7 \cdot 10^{16}$ можливих ключів.

Істотним є також третя риса. Якщо не відома мова відкритого тексту, то можемо не розпізнати результати розшифрування. Більше того, криптограма може бути яким-небудь способом скорочена або спресована, а це додаткові перешкоди на шляху розшифрування. Якщо спресуємо файл, а потім його зашифруємо простим шифром підстановки, то ВТ може бути не розпізнаний цим методом.

Криптовнвліх шифру Цезаря можна ускладнити, якщо поміняти звичний порядок літер в алфавіті. Наприклад, нехай порядок літер змінений таким чином:

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
a	d	g	j	m	p	s	v	y	b	e	h	k	u	q	t	w	z	c	f	i	l	o	r	u	x

Тоді шифрограма буде такою: KMMFYUS OYHH YU FOMHLM і її криптоаналіз стає складнішим.

б) **Блоковий спосіб шифрування** з використанням техніки підстановок. Шифрування слова в $p = y_1 y_2 \dots y_m$ е вищенаведеному алфавіті виконується таким чином:

– слово p розбивається на блоки по t символів в кожному блоці;

– кожний символ в блоці перетворюється за допомогою підстановок $\alpha_1, \alpha_2, \dots, \alpha_t$, кожна з яких має вигляд:

$$\alpha_i(k) = k + j_i \ (mod\ 26),$$

де k – номер літери в алфавіті X, а $k + j_i \ (mod\ 26)$ – її відповідник в алфавіті X при підстановці α_i, $i = 1, 2, \dots, t$.

Тоді, коли $t = 3$, $\alpha_1(k) = k + 3$, $\alpha_2(k) = k + 7$, $\alpha_3(k) = k + 10$, слово $p =$ *thi sci phe ris sec ure* перетворюється до слова

$$q = \textit{wos vjs soo upc vlm xyo}.$$

Дійсно, літера t має номер $k = 19$ і $\alpha_1(k) = 19 + 3 = 22$, а це номер літери w в алфавіті X, літера h має номер $k = 7$ і $\alpha_2(k) = 7 + 7 = 14$, а це номер літери o в алфавіті X, літера i має номер $k = 8$ і $\alpha_3(k) = 8 + 10 = 18$, а це номер літери s в алфавіті X, літера s має номер $k = 18$ і $\alpha_1(k) = 18 + 3 = 21$, а це номер літери v в алфавіті X і т. д.

Дешифрація виконується очевидним чином:

$$\alpha^{-1}(k) = \begin{cases} k - j_i, & \text{якщо } k - j_i \geq 0, \\ k - j_i + 26, & \text{якщо } k - j_i < 0. \end{cases}$$

Наприклад, літери o і c мають таких відповідників:

$\alpha_2^{-1}(14) = 14 - 10 = 4$, а це літера e, $\alpha_3^{-1}(2) = 2 - 10 + 26 = 18$, а це літера s.

Перевагою такого способу шифрування є те, що частота входження літер в текст шифрограми скрита, а це значно ускладнює криптоаналіз такого тексту. ♠.

Абелеві групи. Ці групи отримуємо шляхом додавання закону комутативності до множини тотожностей групи.

Означення 51. *Група $G(A, \Omega)$ називається* **абелевою групою**, *якщо, крім тотожних співвідношень, які визначають її як вільну групу, ця множина включає закон комутативності для операції множення.*

Довільне слово із G можна записати у вигляді

$$x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}, \tag{2.6}$$

де x_j всі попарно різні, n_j – цілі числа, $j = 1, 2, \ldots, k$. Запис слова у вигляді (2.6) називають *мультиплікативним*, але часто для запису слів абелевої групи (як і слів комутативної напівгрупи) використовують адитивний запис – у вигляді суми:

$$n_1 \cdot x_1 + n_2 \cdot x_2 + \ldots + n_k \cdot x_k. \tag{2.7}$$

Числа n_j у цьому виразі називаються **коефіцієнтами**. Додавання слів виду (2.7) визначається як додавання коефіцієнтів при однакових елементах $x_i \in X$, а роль одиниці при такій формі запису відіграє **нуль** – нульовий елемент, тобто вираз виду (2.7), у якого всі коефіцієнти дорівнюють нулю. Якщо деякі x_j із X у виразі (2.7) відсутні, то вважається, що коефіцієнт n_j при цьому x_j дорівнює нулю.

Приклади абелевих груп. 1. Множина $M = \{-1, 1\}$ являє собою групу відносно звичайного множення чисел. Дійсно, $1 \cdot 1 = 1$, $(-1) \cdot 1 = 1 \cdot (-1) = -1$, $(-1) \cdot (-1) = 1$. Отже, множина M замкнута відносно операції множення. Роль одиниці виконує елемент 1. Ця множина також замкнута і відносно операції взяття оберненого (елементи 1 і -1 обернені до самих себе). Очевидно, що операція множення асоціативна і комутативна.

2. Множина $M = \{0, 1\}$ складає абелеву групу відносно операції додавання, якщо покласти 1+1=0 (група лишків за модулем 2). Дійсно, 0+0=0, 0+1=1+0=1, 1+1=0 і, отже, M замкнута відносно операції додавання і взяття оберненого (1 є елементом оберненим до самого себе). Одиничним елементом служить 0. Очевидно, що операція додавання асоціативно-комутативна.

Дану конструкцію можна узагальнити. Нехай $Z_n = \{0, 1, \ldots, n-1\}$, $n > 1$. Задамо на множині Z_n операцію додавання $\oplus$:

$$k \oplus l = \begin{cases} k + l, & \text{якщо} \quad k + l < n; \\ k + l - n, & \text{якщо} \quad k + l \geq n, \end{cases}$$

де $+$ і $-$ звичайні операції додавання і віднімання цілих чисел, $k, l \in Z_n$. Множина Z_n з заданою таким чином операцією додавання складає абелеву групу n-го порядку. Дійсно, роль нуля відіграє елемент 0, роль оберненого елемента до даного елемента k – елемент l такий, що $k + l = n$. Очевидно, що операція додавання комутативна на підставі комутативності операції додавання цілих чисел. Доведення асоціативності операції пропонується як проста вправа.

Група Z_n називається **групою лишків за модулем** n і виникає в результаті розбиття множини цілих чисел $\mathcal{Z}$ на класи, де в один клас попадають ті і тільки ті числа, які при діленні на $n > 1$ дають однакові остачі. Якщо позначити $rest(m, n)$ остачу від ділення числа m на n, то операцію додавання $\oplus$ можна визначити іншим способом:

$$k \oplus l = rest(k + l, n).$$

Очевидно, що розглянута вище група $M = \{0, 1\}$ збігається з групою Z_2.

3. Множина цілих чисел $\mathcal{Z}$ складає адитивну групу, але $\mathcal{Z}$ не є групою відносно операції множення, оскільки операція ділення в множині $\mathcal{Z}$ (операція взяття оберненого) не завжди визначена.

Множина Z_O цілих парних чисел складає адитивну групу і є підгрупою групи $\mathcal{Z}$. Взагалі адитивною підгрупою буде довільна множина цілих чисел, які кратні деякому заданому цілому числу n.

Множина непарних чисел не буде групою, оскільки вона не замкнута відносно операції додавання.

4. Циклічна група G, тобто група, породжена одним елементом, наприклад a, складається із елементів виду a^n, де $n \in \mathcal{Z}$, $a^0 = e, a^1 = a$. Множення елементів визначається як додавання степенів, тобто $a^k \cdot a^l = a^{k+l}$.

5. Скінченна триелементна група поворотів правильного трикутника на 0, 120, 240 градусів. Ці повороти є такими, які суміщують трикутник сам з собою. Добутком двох поворотів називається поворот, отриманий в результаті послідовного застосування першого з поворотів, а потім другого. Якщо позначити поворот на 0 градусів через a_0, поворот на 120 градусів – a_1 і поворот на 240 градусів – a_3, то отримуємо таку таблицю Келі для даної алгебри:

$*$	a_0	a_1	a_2
a_0	a_0	a_1	a_2
a_1	a_1	a_2	a_0
a_2	a_2	a_0	a_1

З цієї таблиці очевидним чином випливає, що поворот на 0 градусів відіграє роль одиничного елемента групи, ця група комутативна, оскільки таблиця симетрична. ♠

Теорема 13. *Довільна нескінченна циклічна група ізоморфна адитивній групі цілих чисел Z. Довільна скінченна циклічна група n-го порядку ізоморфна групі лишків за модулем n – Z_n.*

Доведення. У першому випадку відображення h задається $h(a^n) = n$, а в другому $h(a^k) = rest(k, n)$.
Покажемо, що h – ізоморфізм для першого випадку.
a1) $h(a^m \cdot a^n) = h(a^{m+n}) = m + n$;
a2) $h(a^{-n}) = -n$,
і оскільки a^n обернений до a^{-n}, то умови ізоморфізму виконуються;
a3) $h(a^0) = h(e) = 0$.
Залишається показати, що h – взаємно однозначне. Для цього необхідно зауважити, що із $a^m = a^n$ випливає $a^{m-n} = a^0 = e$, тобто $m = n$. ■

Другий випадок пропонується розглянути як вправу. ♠

2.5.3. Побудова "нетрадиційної" арифметики на основі абелевих груп

Нехай задана деяка скінченна множина цілих чисел, наприклад, $N_5 = \{0, 1, 2, 3, 4\}$. Оскільки ми хочемо побудувати адитивну абелеву групу, то ця множина обов'язково повинна включати 0. Для того, щоб N_5 перетворити в групу GN_5, необхідно коректно задати значення для операції додавання з одним із елементів групи, скажімо з 1. Дійсно, оскільки $a+0 = a$ для довільного $a \in GN_5$, то перший рядок таблиці додавання елементів групи визначений (таблиця 1), а на підставі комутативності (оскільки GN_5 абелева) і перший стовпчик цієї таблиці. Нехай, наприклад, задано $0 + 1 = 1$, $1 + 1 = 4$, $1 + 4 = 2$, $1 + 2 = 3$, $1 + 3 = 0$. Таке задання коректне, оскільки має місце єдиність результату (але єдиність результату, як буде показано нижче, не достатня умова ґарантії коректності). Тепер послідовно знаходимо:

$$2{+}2{=}2{+}(1{+}4){=}(2{+}1){+}4{=}3{+}(1{+}1){=}(3{+}1){+}1{=}0{+}1{=}1,$$

$2{+}3{=}2{+}(1{+}2){=}(2{+}2){+}1{=}1{+}1{=}4$, $\quad 2{+}4{=}2{+}(1{+}1){=}(2{+}1){+}1{=}3{+}1{=}0$,
$3{+}3{=}3{+}(1{+}2){=}(3{+}1){+}2{=}0{+}2{=}2$, $\quad 3{+}4{=}3{+}(1{+}1){=}(3{+}1){+}1{=}0{+}1{=}1$,
$4{+}4{=}4{+}(1{+}1){=}(4{+}1){+}1{=}2{+}1{=}3$.

Заносимо ці значення в таблицю 2 і на цьому закінчуємо побудову групи GN_5.

Таблиця 1

+	0	1	2	3	4
0	0	1	2	3	4
1	1	4	3	0	2
2	2	3			
3	3	0			
4	4	2			

Таблиця 2

+	0	1	2	3	4
0	0	1	2	3	4
1	1	4	3	0	2
2	2	3	1	4	0
3	3	0	4	2	1
4	4	2	0	1	3

Аналогічно можна задати і довільну іншу групу GN_5, яка включає 0. Дійсно, якщо, наприклад, $M_5 = \{0, 2, 6, 3, 5\}$, то встановимо взаємно однозначну відповідність h між N_5 і M_5 (ця відповідність існує, оскільки N_5 і M_5 рівнопотужні), яка переводить 0 в 0, а останні елементи – довільним чином. Наприклад, нехай задана відповідність

$$1 \leftrightarrow 6, 2 \leftrightarrow 3, 3 \leftrightarrow 5, 4 \leftrightarrow 2, 0 \leftrightarrow 0.$$

Звідси отримуємо таку таблицю.

Таблиця 3

+	0	2	6	3	5
0	0	2	6	3	5
2	2	5	3	0	6
6	6	3	2	5	0
3	3	0	5	6	2
5	5	6	0	2	3

Перевіримо, наприклад, чому відповідає в цій таблиці $4 = 1{+}1$. Маємо $h(1+1) = h(4) = h(1) + h(1) = 6 + 6 = 2$.

В якості вправи, пропонується виконання перевірки для всієї таблиці 3.

Зауважимо, що для побудови групи GN_k, мало вимагати тільки однозначності операції додавання. Якщо визначити додавання в групі так, що в ній буде елемент скінченного порядку (тобто група з крученням), то коректність визначення операції може бути порушена. Наприклад, якщо задати додавання

$$0+1=1,\ 1+1=0,\ 1+2=3,\ 1+3=4,\ 1+4=2,$$

то, виконуючи перевірку (обчислюючи) $1{+}3$, отримаємо

$$1+3=1+(1+2)=(1+1)+2=0+2=2,$$

що не співпадає з визначеним вище. Справа в тім, що елемент 1 є елементом 2-го порядку $(1+1=0)$, і це вносить свої корективи при визначенні операції (її вже не можна задавати довільним чином).

Далі будуть розглянуті інші застосування груп, зокрема, при розв'язанні проблеми передачі ключів.

2.5.4. Кільця

Ця алгебра будується шляхом розширення сигнатури операцій та множини тотожних співвідношень для цих операцій.

Означення 52. *Універсальна алгебра $G(A,\Omega)$ називається* **кільцем**, *якщо вона є*

а) *абелевою групою відносно додавання;*

б) *групоїдом відносно множення;*

в) *і включає закони дистрибутивності, тобто для довільних її елементів x, x', x''*

$$x(x'+x'') = (xx') + (xx''), \qquad (x+x')x'' = (xx'') + (x'x'').$$

Це означає, що Ω включає чотири операції: бінарні операції додавання і множення, унарну операцію взяття оберненого відносно операції додавання і нульарну операцію, яка фіксує нульовий елемент абелевої групи кільця. Цей нульовий елемент називається **нулем** кільця.

Множення у кільці зводиться, на підставі законів дистрибутивності, до множення елементів із X, яке виконується за правилом множення слів у групоїді.

Із законів 1) – 3) випливають співвідношення, які дає

Твердження 2. *У довільному кільці K для довільних його елементів a, b, c має місце:*

а) $a(b-c) = ab - ac$, $(b-c)a = ba - ca$;

б) $a0 = 0a = 0$; в) $(-a)b = a(-b) = -ab$, $(-a)(-b) = ab$.

Доведення. а) На підставі комутативності й асоціативності операції додавання можна записати $c + (b - c) = b$. Домноживши обидві частини цього рівняння зліва на a, дістаємо

$$a(c + (b - c)) = ac + a(b - c) = ab.$$

Звідси, $(ac+a(b-c))-ac = ab-ac = a(b-c)$, знову ж таки на підставі комутативності й асоціативності операції додавання.

Аналогічно доводиться і другий закон.

б) Нехай b – довільний елемент кільця. Тоді, на підставі доведеного вище пункту а), маємо

$$a0 = a(b - b) = ab - ab = 0.$$

в) *Доведення* пропонується як вправа.

Зауважимо, що обернене твердження до пункту б) в довільному кільці не має місця, тобто існують такі кільця, в яких є відмінні від нуля елементи a, b, добуток яких дорівнює нулю ($ab = 0$). Якщо такі елементи в кільці є, то вони називаються **дільниками нуля**.

Означення 53. *Кільце називається* **асоціативним (комутативним)**, *якщо його операція множення асоціативна (комутативна), і називається* **кільцем з одиницею**, *коли воно має одиничний елемент відносно операції множення.*

Кільце називається **асоціативно-комутативним**, *якщо воно асоціативне і комутативне одночасно.*

Множення у асоціативному кільці виконується за правилом множення слів у напівгрупі, а сама напівгрупа називається **мультиплікативною напівгрупою** асоціативного кільця.

Елементи асоціативно-комутативного кільця з множиною вільних твірних X являють собою многочлени від елементів із X з цілими коефіцієнтами. Тому таке кільце часто називають просто **кільцем многочленів над X**.

Приклади кілець. 1. Прикладом асоціативного кільця з одиницею може служити множина квадратних матриць над довільним кільцем P з одиницею.

Нехай $M(p, q. P)$ – множина всіх матриць розмірності $p \times q$ над P, а $M(p, P)$ – множина квадратних матриць $M(p, p, P)$.

Матриця C, яка складається з елементів виду $a_{ij} + b_{ij}$, де a_{ij}, b_{ij} – відповідно елементи матриць A B із $M(p, q, P)$, називається **сумою матриць** A і B. Оскільки множина P є кільцем, то множина $M(p, q, P)$, очевидно, є абелевою групою відносно додавання. Роль нульового елемента відіграє матриця, у якої всі елементи складаються із нулів (*нульова матриця*), а роль оберненої матриці для матриці A відіграє матриця A', у якої $a'_{ij} = -a_{ij}$, де a'_{ij}, a_{ij} – елементи матриць A' і A відповідно.

Нехай $A \in M(p, q, P)$, $B \in M(q, r, P)$, тоді матриця C, яка складається з елементів

$$c_{ij} = \sum_{k=1}^{q} a_{ik} \cdot b_{kj}$$

називається **добутком матриць** A і B. З цього означення випливає, що не довільні дві матриці можна перемножити, а лише *відповідні*, тобто такі, у яких число стовпчиків першої дорівнює числу рядків другої. Ясно, що квадратні матриці із $M(p, P)$ відповідні, і для таких матриць добуток завжди визначений.

Розглянемо множину матриць $M(p, P)$. Матриця A називається **одиничною**, якщо $a_{ij} = 0$ при $i \neq j$ і $a_{ii} = 1_P$, де 1_P – одиниця кільця P. Одинична матриця, як правило, позначається через E. Легко впевнитися, що $\forall A \in M(p, P)$ має місце рівність $A \cdot E = E \cdot A = A$, тобто матриця E відіграє роль одиничного елемента в множині $M(p, P)$.

Неважко показати, що $M(p, P)$ є асоціативним кільцем з одиницею. Покажемо, наприклад, справедливість одного з дистрибутивних законів:

$$(\forall A, B \in M(p, P))\ A \cdot (B + C) = A \cdot B + A \cdot C.$$

Дійсно, за визначенням добутку матриць маємо

$$\sum_{k=1}^{p} a_{ik}(b_{kj} + c_{kj}) = \sum_{k=1}^{p} (a_{ik} \cdot b_{kj} + a_{ik} \cdot c_{kj}) = \sum_{k=1}^{p} a_{ik} \cdot b_{kj} + \sum_{k=1}^{p} a_{ik} \cdot c_{kj},$$

тобто $A \cdot (B + C)$ і $A \cdot B + A \cdot C$ мають однакові елементи.

2. Множина $M = \{0, 1\}$ буде кільцем, якщо визначити додавання і множення елементів цієї множини таким чином:

$$0 \oplus 0 = 1 \oplus 1 = 0, 0 \oplus 1 = 1 \oplus 0 = 1, 0 \odot 1 = 1 \odot 0 = 0, 1 \odot 1 = 1.$$

Неважко перевірити, що дана множина буде асоціативно-комутативним кільцем з одиницею, яке називається *кільцем лишків за модулем числа 2* і позначається Z_2.

Цю конструкцію можна узагальнити і отримати **кільце лишків Z_m за модулем** числа m. Дійсно, нехай $m \neq 0$ – фіксоване ціле число. Два числа a і b називаються такими, що порівнюються за модулем m, якщо вони при діленні на m дають однакові остачі. Відношення порівнюваності за модулем m розбиває множину цілих чисел на класи, які позначають числами $0, 1, \ldots, m-1$ відповідно до остач, які виникають при виконанні ділення.

Визначивши операції додавання і множення так, що $\forall x, y \in \{0, 1, \ldots, m-1\}$

$$x \oplus y = \begin{cases} x + y & \text{якщо } x + y \leq m, \\ x + y - n & \text{якщо } x + y > m, \end{cases} \qquad x \odot y = \begin{cases} xy & \text{якщо } xy \leq m, \\ rest(xy, n) & \text{якщо } xy > m \end{cases}$$

дістаємо кільце лишків з одиницею за модулем m. Нулем кільця служить 0, одиницею – 1.

Наприклад, якщо $m = 6$, то таблиці введених операцій такого кільця набувають вигляду:

$\oplus$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

,

$\odot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Як видно з таблиці множення, це кільце має дільники нуля 2 і 3 та 3 і 4.

Детальніше властивості цих кілець будуть розглянуті в наступних розділах.

3. Асоціативно-комутативне кільце називається **булевим**, якщо для довільного елемента a цього кільця має місце співвідношення $a^2 = a$. Ця тотожність називається **законом ідемпотентності**.

Твердження 3. *У довільному булевому кільці для довільних елементів a, b мають місце співвідношення*

$$\text{а) } 2 \cdot a = 0; \qquad \text{б) } a \cdot b = b \cdot a.$$

Доведення. а) На підставі ідемпотентності $a+b = (a+b)^2 = a+a \cdot b+b \cdot a+b$. Звідси отримуємо $a \cdot b+b \cdot a = 0$. Покладемо $b = a$, тоді $a^2 + a^2 = a + a = 2 \cdot a = 0$. Звідси знаходимо також, що $a = -a$.

б) Із (а) одержуємо $a \cdot b = b \cdot a = 0$. Міняючи a на $-a$ в одному із доданків, знаходимо $a \cdot b - b \cdot a = 0$ або $a \cdot b = b \cdot a$. ■ ♠

Асоціативно-комутативне кільце без дільників нуля називається **областю цілісності**.

Областями цілісності, як легко переконатися, будуть кільця цілих і раціональних чисел – $\mathcal{Z}$ і RC. Областю цілісності буде також і кільце многочленів.

Дійсно, нехай KC – довільне асоціативно-комутативне кільце. Розглянемо всі можливі многочлени вигляду

$$a_0 + a_1 \cdot x_1^1 + a_2 \cdot x_1^2 + \ldots + a_n \cdot x_1^n,$$

де $n_i \in \mathcal{N}$, відносно невідомого x_1 з коефіцієнтами a_i із RC, $i = = 1, \ldots, n$. Якщо $a_n \neq 0$, то число n називається *степенем* цього многочлена. Визначаючи додавання і множення многочленів звичайним шляхом так, як це прийнято в курсі вищої алгебри, отримуємо **кільце многочленів** $R[x_1]$ від невідомого x_1 над кільцем KC. Нулем кільця $R[x_1]$ служить многочлен, всі коефіцієнти якого дорівнюють нулю.

Аналогічно можна визначити кільце многочленів $R[x_1, \ldots, x_n]$ від довільного скінченного числа невідомих як кільце многочленів від одного невідомого x_n над кільцем $R[x_1, \ldots, x_{n-1}]$. Має місце

Теорема 14. *Якщо KC область цілісності, то $R[x_1, x_2, \ldots, x_n]$ теж область цілісності.*

Вище було визначено поняття дільника нуля в кільці. Подібно до поняття дільника нуля існує і поняття дільника одиниці.

Означення 54. *Нехай G – довільне кільце і $a, b \in G$. Елемент a називається дільником b або "a ділить b" або "b кратне a", якщо рівняння $ax{=}b$ має хоча б один розв'язок.*

Елемент $a \in G$ називається дільником одиниці, якщо він ділить всі елементи кільця G.

Термін дільник одиниці має таке пояснення: якщо рівняння $ax = 1$ має розв'язок, то рівняння $ay = b$ буде мати розв'язок для довільного $b \in G$. Дійсно, якщо $x = c$ – розв'язок рівняння $ax = 1$, то $y = cb$ буде розв'язком рівняння $ay = b$ (що перевіряється простою підстановкою).

Неважко показати, що множина всіх дільників одиниці утворює абелеву групу відносно операції множення (див. вправи в кінці розділу).

Ідеали кільця. Поняття, аналогічне в певному сенсі поняттю нормального дільника групи, в кільцях є ідеал.

Означення 55. *Підмножина I кільця K називається ідеалом кільця K, якщо*
а) $\forall a, b \in I$ $a \pm b \in I$;
б) якщо $x \in K$ і $r \in I$, то $xr, rx \in I$.

Нехай K – комутативне кільце і I – його ідеал. Ідеал I називається головним ідеалом, породженим елементом a (позначення $I(a)$), якщо $I(a) = \{ar | r \in K\}$.

Прикладами ідеалів кільця цілих чисел $\mathcal{Z}$ є множина всіх цілих чисел, кратних фіксованому числу k. Довільний ідеал кільця цілих чисел $\mathcal{Z}$ буде головним, як випливає з наступної теореми.

Теорема 15. *Кожний непустий ідеал кільця цілих чисел $\mathcal{Z}$ – головний ідеал.*

Доведення. Для $I \neq \{0\}$ нехай k – найменше додатне число, яке належить ідеалу I, і нехай $m \in I$ – довільне число із $\mathcal{Z}$. Тоді $m = kn + r$, де $0 \leq r < k$. Оскільки $r = m - kn \in I$, то це можливо тільки тоді, коли $r = 0$ на підставі того, що k вибиралося найменшим. А це означає, що $m = kn$, тобто m кратне k. ■

Приклад ідеалів. Нехай $\mathcal{Z}$ – кільце цілих чисел. Розглянемо два ідеали цього кільця $I(8)$ та $I(12)$:
$I(8) = \{8r | r \in \mathcal{Z}\} = \{\ldots, -24, -16, -8, 0, 8, 16, 24, \ldots\}$;
$I(12) = \{12s | s \in \mathcal{Z}\} = \{\ldots, -24, -12, 0, 12, 24, \ldots\}$.

Перетин цих ідеалів $I(8) \cap I(12) = \{\ldots, -48, -24, 0, 24, 48, \ldots\}$ теж є головним, оскільки він породжується цілим число 24, а це число дорівнює найменшому спільному кратному чисел 8 і 12. ♠

Теорема 16. *Якщо r і s ненульові цілі числа, а $I(s)$ і $I(r)$ – відповідні головні ідеали кільця $\mathcal{Z}$, то*

а) якщо $I(s) \subseteq I(r)$, то $r|s$;

б) $I(s) \cap I(r) = I(НСК(s, r))$;

в) якщо $I(s,r)$ – найменший ідеал, який включає числа s і r, то $I(s, r) = I(НСД(s, r))$.

Доведення. Нехай $m \in I(s)$, тоді $m \in I(R)$ і тому $m = sk = rk_1$, де k, k_1– довільні цілі числа. Отже, sk повинно бути кратним r. Але на підставі того, що k довільне, то s повинно ділитися на r.

б) Нехай $m \in I(s) \cap I(r)$, тоді $m = srk$, де k – довільне ціле число із $\mathcal{Z}$. Отже, число sr породжує ідеал $I(s) \cap I(r)$ і найменшим таким числом буде найменше спільне кратне чисел s і r.

в) Доведення пропонується як вправа. ■

Приклад 3.2.1. Нехай $\mathcal{Z}$ – кільце цілих чисел. Розглянемо два ідеали цього кільця $I(8)$ та $I(12)$. Тоді, ідеал, який включає числа 8 і 12 має вигляд:

$I(8, 12) = I(4) = \{\ldots, -8, -4, 0, 4, 8, \ldots\}$. ♠

2.5.5. Приклад застосування кілець в криптографії

Розглядаючи абелеві групи, був побудовний приклад "нетрадиційної арифметики" для операції додавання. Для зручності наведемо таблицю додавання абелевої групи GN_5, яка була побудована вище.

+	0	1	2	3	4
0	0	1	2	3	4
1	1	4	3	0	2
2	2	3	1	4	0
3	3	0	4	2	1
4	4	2	0	1	3

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	3	4	1
3	0	3	4	1	2
4	0	4	1	2	3

Кільця дозволяють розширити таку побудову і на операцію множення (таблиця справа).

Розглянемо як довизначається група GN_5 до кільця з одиницею. Роль одиниці буде відігравати 1. На підставі аксіом кільця і твердження 2 маємо: для довільного елемента a із GN_5 $a \cdot 1 = 1 \cdot a = a$, $a \cdot 0 = 0 \cdot a = 0$. Таким чином, два рядки і два стовпчики таблиці множення вже визначені. Далі, використовуючи таблицю додавання і закон дистрибутивності, дітстаємо таблицю для операції множення. Дійсно,

$$4 \cdot 2 = (1+1) \cdot 2 = 2+2 = 1;\ 4 \cdot 3 = (1+1) \cdot 3 = 3+3 = 2;\ 4 \cdot 4 = (1+1) \cdot 4 = 4+4 = 3.$$

Легко бачити, що $4 \cdot 2 = 2 \cdot 4$ і $4 \cdot 3 = 3 \cdot 4$, тобто елемент 4 комутативний зі всіма останніми. Далі

$$2 \cdot 2 = (1+4) \cdot 2 = 2+4 \cdot 2 = 2+1 = 3;\ 2 \cdot 3 = (1+4) \cdot 3 = 3+4 \cdot 3 = 3+2 = 4;$$
$$2 \cdot 4 = (1+4) \cdot 4 = 4+4 \cdot 4 = 4+3 = 1.$$

Аналогічно одержуємо $3 \cdot 3 = (2+1) \cdot 3 = 2 \cdot 3 + 3 = 4+3 = 1$ решту елементів таблиці 4. Із симетричності таблиці випливає, що множина елементів GN_5 комутативна. Крім того, легко перевірити, що GN_5 також і асоціативна, тобто GN_5 – асоціативно-комутативне кільце з одиницею.

Розглянемо тепер приклад кільця без одиниці. Візьмемо групу GM5, операція додавання якої задана таблицею 3:

Таблиця 3

+	0	2	6	3	5
0	0	2	6	3	5
2	2	5	3	0	6
6	6	3	2	5	0
3	3	0	5	6	2
5	5	6	0	2	3

Для того, щоб перетворити цю групу в кільце, необхідно, як і у випадку додавання, задати хоча б один рядок для операції множення.

Нехай задано: $2 \cdot 2 = 3,\ 2 \cdot 3 = 5,\ 2 \cdot 5 = 6,\ 2 \cdot 6 = 2$. Тоді маємо:

$$5 \cdot 2 = (2+2) \cdot 2 = 2 \cdot 2 + 2 \cdot 2 = 3+3 = 6; 5 \cdot 3 = (2+2) \cdot 3 = 2 \cdot 3 + 2 \cdot 3 = 5+5 = 3;$$
$$5 \cdot 6 = (2+2) \cdot 6 = 2 \cdot 6 + 2 \cdot 6 = 2+2 = 5; 5 \cdot 5 = (2+2) \cdot 5 = 2 \cdot 5 + 2 \cdot 5 = 6+6 = 2;$$
$$6 \cdot 2 = (5+2) \cdot 2 = 5 \cdot 2 + 2 \cdot 2 = 6+3 = 5; 6 \cdot 3 = (5+2) \cdot 3 = 5 \cdot 3 + 2 \cdot 3 = 3+5 = 2;$$
$$6 \cdot 5 = (5+2) \cdot 5 = 5 \cdot 5 + 2 \cdot 5 = 2+6 = 3; 6 \cdot 6 = (5+2) \cdot 6 = 5 \cdot 6 + 2 \cdot 6 = 5+2 = 6;$$
$$3 \cdot 2 = (6+2) \cdot 2 = 6 \cdot 2 + 2 \cdot 2 = 5+3 = 2; 3 \cdot 3 = (6+2) \cdot 3 = 6 \cdot 3 + 2 \cdot 3 = 2+5 = 6;$$
$$3 \cdot 6 = (6+2) \cdot 6 = 6 \cdot 6 + 2 \cdot 6 = 6+2 = 3; 3 \cdot 5 = (6+2) \cdot 5 = 6 \cdot 5 + 2 \cdot 5 = 3+6 = 5.$$

тобто, одержуємо таку таблицю множення:

Таблиця 4

·	0	2	6	3	5
0	0	0	0	0	0
2	0	3	2	5	6
6	0	5	6	2	3
3	0	2	3	6	5
5	0	6	5	3	2

З цієї таблиці видно, що кільце некомутативне. Легко впевнитися, що воно і неасоціативне, наприклад $(3 \cdot 6) \cdot 5 = 3 \cdot 5 = 5$ і $3 \cdot (6 \cdot 5) = 3 \cdot 3 = 6$ не рівні між собою.

Зауважимо, що для того, щоб оптимальним шляхом знаходити таблицю як для додавання, так і для множення слід поступати так: якщо визначено множення на число 2 (як в попередньому прикладі), то в таблиці додавання знаходимо суму 2+2 (вона рівна 5) і беремо результат цієї суми як перший множник. Далі поступаємо аналогічно для числа 2 і числа, яке отримане раніше (в прикладі це число 5), шукаємо добуток цих чисел. Знаходимо результат множення для них і т. д., доки не побудуємо всі елементи таблиці.

При обчисленнях на комп'ютерах така методика часто відіграє важливу роль при побудові ефективних алгоритмів.

Зі сказаного випливає такий спосіб шифрування з використанням властивостей кільця:

1) Будуємо абелеву групу порядок якої $n \geq 26$;

2) Визначаємо операцію додавання $1 + 1 = a$ з метою побудови кільця з одиницею (елемент a служить ключем) або додатково визначаємо операцію множення $a \cdot a = b$ (елемент b служить ключем) з метою побудови кільця без одиниці:

3) Добудовуємо таблиці додавання і множення за визначеними значеннями в 2);

4) Задаючи бієкцію f між елементами кільця (або для таблиці додавання, або для таблиці множення) і літерами алфавіту, шифруємо текст. Шифрування можна виконувати з використанням однієї з таблиць, або з використанням обох таблиць з повторним шифруванням.

5) Дешифрація виконується в зворотному порядку: знаходимо значення f^{-1} на символах шифрограми; потім на основі таблиць кільця повністю дешифруємо отриманий текст.

2.5.6. Поля

Означення 56. *Асоціативно-комутативне кільце з одиницею називається* **полем**, *якщо воно як відносно додавання, так і відносно множення є абелевою групою. Група поля відносно додавання називається* **адитивною**, *а група відносно множення* – **мультиплікативною**.

Зауважимо, що поле є областю цілісності і, як це випливає із теореми 14, кільце многочленів над довільним полем P є областю цілісності. Одиницею в цій області цілісності служить многочлен, коефіцієнти якого всі рівні нулю, крім коефіцієнта a_0, який дорівнює 1_P – одиниці поля P.

Поле $F = (A, \Omega)$ називається скінченним, якщо його носій A має скінченни число елементів.

Нехай $a \in F$, тоді елементи $a, a + a, a + a + a, \ldots$ є елементами поля, які будемо позначати $a, 2a, 3a, \ldots, na, \ldots$ відповідно (елемент n не обов'язково повинен бути елементом поля F). Аналогічно елементи $a, a \cdot a, a \cdot a \cdot a, \ldots$ теж є елементами поля і їх позначають $a, a^2, a^3, \ldots, a^n, \ldots$ відповідно. Нехай $a \neq 0$ є елементом поля F.

Означення 57. *Якщо існує таке найменше число $n \in \mathcal{N}$, що $n \cdot a = 0$, то n називається адитивним порядком елемента a. Якщо існує таке найменше число $n \in \mathcal{N}$, що $a^n = 1$, то n називається мультиплікативним порядком елемента a.*

Теорема 17. *Всі ненульові елементи поля F мають один і той же адитивний порядок.*

Доведення. Нехай $a, b \in F \setminus \{0\}$ і їх адитивні порядки дорівнюють n і m відповідно. Тоді

$$n \cdot b = n \cdot (a \cdot a^{-1}) \cdot b = (n \cdot a) \cdot (a^{-1} \cdot b) = 0 \cdot a^{-1} \cdot b = 0.$$

Звідси випливає, що $m \leq n$. Аналогічно

$$m \cdot a = m \cdot (b \cdot b^{-1}) \cdot a = (m \cdot b) \cdot (b^{-1} \cdot a) = 0 \cdot b^{-1} \cdot a = 0.$$

Звідси випливає, що $n \leq m$ і тому $m = n$. ■

Означення 58. *Якщо в полі F всі ненульові елементи мають адитивний порядок n, то поле F називається полем характеристики n. Якщо такого числа n не існує, то поле називається полем характеристики 0.*

Теорема 18. *Характеристика довільного скінченного поля є простим числом.*

Доведення. Нехай F – скінченне поле F і $a \in F \setminus \{0\}$ – довільний його елемент. Тоді в послідовності $a, 2a, 3a, \ldots$ існують такі числа $i, j \in \mathcal{Z}, i < j$, що $ja = ia$ або $(j - i)a = 0$. Отже, поле F має додатну характеристику, яку позначимо n. На підставі того, що поле F включає принаймні два елементи (0 і 1), то $n \geq 2$. Якщо число n не є простим, то існують такі числа $k, l \in \mathcal{Z}, 1 < k, l < n$, що $n = kl$. Тоді

$$0 = n \cdot 1 = (k \cdot l) \cdot 1 = (k \cdot l) \cdot (1 \cdot 1) = (k \cdot 1)(l \cdot 1).$$

Оскільки поле є областю цілісності, то або $k \cdot 1 = 0$ або $l \cdot 1 = 0$. Звідси дістаємо, що або $ka1 = (k1)a = 0$ або $la1 = (l1)a = 0$ для всіх $a \in F$. Але це суперечить означенню характеристики поля n. ■

З доведеної теореми випливає такий основний результат для скінченних полів.

Теорема 19. *Скінченне поле F має характеристику p і порядок p^n для деякого $n \in \mathcal{N}$.*

Доведення. З попередньої теореми випливає, що поле F має характеристику p, де p – просте число. Нехай $|F| = q$. Якщо $q = p$, то твердження теореми, очевидно, має місце. В протилежному випадку візьмемо елемент $a_1 \in F \setminus \{0\}$ і покладемо

$$G_1 = \{y : y = m_1 \cdot a_1,\ m_1 \in \mathcal{N},\ 1 \leq m_1 \leq p\}.$$

Розглянемо тепер елемент $a_2 \in F \setminus G_1$ і побудуємо

$$G_2 = \{z : z = m_1 \cdot a_1 + m_2 \cdot a_2,\ m_1, m_2 \in \mathcal{N},\ 1 \leq m_1, m_2 \leq p\}.$$

Якщо $F \neq G_2$, то розглядається елемент $a_3 \in F \setminus G_2$ і т. д. На підставі скінченності поля F такий процес закінчується і ми отримуємо сукупність множин $G_1, G_2, \ldots, G_n$ для деякого $n \in \mathcal{N}$.

Кожний елемент $a \in F$ єдиним чином представляється у вигляді

$$a = m_1 \cdot a_1 + m_2 \cdot a_2 + \ldots + m_n \cdot a_n,$$

де $1 \leq m_i \leq p$, для всіх $i = 1, 2, \ldots, n$. (Довести це як корисну вправу). Отже, існує p^n таких виразів і тому $|F| = p^n$. ■

Означення 59. *Поле називається простим, якщо воно не має жодного власного підполя.*

Очевидно, що коли поле має простий порядок, то воно є простим полем і порядок простого поля дорівнює характеристиці цього поля. Як випливає з попереднього твердження, простими полями не вичерпуються скінченні поля, оскільки для довільних простого числа p і натурального числа n існує поле порядок якого p^n. Цей більш загальний вигляд скінченних полів будується за допомогою поліномів.

Нехай $F = (A, \Omega)$ – поле. Нагадаємо, що поліномом над полем називається вираз

$$f(x) = \sum_{i=0}^{n} a_i x^i,$$

де $n \in \mathcal{N}$, $a_i \in A$, $0 \leq i \leq n$, а x – символ, який не належить полю F. Коефіцієнт $a_n \neq 0$ називається старшим, якщо $n \neq 0$. Число n називається степенем полінома $f(x)$ і позначається $n = deg(f)$. Якщо старшим коефіцієнтом є a_0, то поліном $f(x)$ називається константою, а коли старший коефіцієнт $a_0 = 0$, то поліном $f(x)$ називається нульовим $f(x) = 0$. Множину всіх поліномів над полем F позначатимемо $F(x)$.

Якщо $f(x), g(x) \in F(x)$, де

$$f(x) = \sum_{i=0}^{n} a_i x^i, \qquad g(x) = \sum_{j=0}^{m} b_j x^j,$$

то

$$f(x) + g(x) = \sum_{k=0}^{\max(m,n)} c_k x^k, \tag{2.8}$$

де

$$c_k = \begin{cases} a_k + b_k, & k = 0, 1, \ldots, \min(m, n), \\ a_k, & k = m+1, \ldots, n, \text{ якщо } m < n, \\ b_k, & k = n+1, \ldots, m, \text{ якщо } n < m. \end{cases}$$

і

$$f(x) \cdot g(x) = \sum_{k=0}^{m+n} c_k x^k, \tag{2.9}$$

де $c_k = \sum\limits_{i+j=k,\ 0 \le i,j \le n}^{m+n} a_i b_j$.

Означення 60. *Поліном $f(x) \in F(x)$ називається простим або незвідним над полем F, якщо він має додатний степінь і з рівності $f(x) = g(x) \cdot h(x)$, де $g(x), h(x) \in F(x)$, випливає що або поліном $g(x)$ є константою, або поліном $h(x)$ є константою. В протилежному випадку поліном називається непростим.*

Побудуємо тепер скінченне поле за допомогою простого полінома. Зауважимо, що для поліномів $f(x)$ і $g(x)$, де $g(x) \neq 0$ як і при діленні цілих чисел, має місце розклад

$$f(x) = g(x) \cdot h(x) + r(x), \tag{2.10}$$

де $g(x), r(x) \in F(x)$ і $deg(r) < deg(g)$.

Означення 61. *Нехай $f(x), g(x), h(x) \in F(x)$, де $g(x) \neq 0$, задовольняють умові (2.10). Тоді поліном $r(x)$ називається остачею від ділення полінома $f(x)$ на поліном $g(x)$. Цей поліном позначається як $r = f \pmod g$. Остачі від ділення всіх поліномів із множини $F(x)$ за модулем полінома $g(x)$ називаються поліномами із множини $F(x)$ за модулем полінома $g(x)$. Множина всіх таких поліномів позначимо $F_g(x)$.*

Очевидно, що степені всіх поліномів із $F_g(x)$ менші $deg(g)$.

Теорема 20. *Нехай F – поле, а $f(x)$ – ненульовий поліном із $F(x)$. Тоді $F_f(x)$ буде полем тоді і тільки тоді, коли f простий над полем $F(x)$.*

Доведення. Спочатку зазначимо, що $F_f(x)$ є кільцем з нулем та одиницею відносно операцій додавання та множення за модулем полінома $f(x)$. Це випливає із співвідношень (2.8), (2.9) і (2.10). Нулем і одиницею виступають 0 і 1 поля F.

Припустимо, що $F_f(x)$ є полем і $f = g \cdot h$, де g і h – поліноми з множини $F(x)$, які не є константами. Тоді, оскільки $0 < deg(g) < < deg(f)$ і $0 < deg(h) < deg(f)$, то поліноми g і h не є константами в множині $F_f(x)$, не дивлячись на те, що поліном f є нульовим в полі $F_f(x)$. Але це суперечить замкнутості операції множення в мультиплікативній групі поля $F_f(x)$. Отже, $F_f(x)$ не може бути полем, а це суперечить простоті полінома $f(x)$ над полем F.

Нехай поліном $f(x)$ є простим над полем F. Оскільки $F_f(x)$ кільце, то для доведення теореми потрібно показати, що для довільного ненульового полінома із $F_f(x)$ в ньому існує елемент, обернений відносно операції множення. Нехай r – ненульовий

поліном із $F_f(x)$ такий, що НСД$(f,r) = c$. Оскільки $f(x)$ простий пліном над полем F і $deg(r) < deg(f)$, то поліном c повинен бути константою. Розглянемо поліном $h(x) = c \cdot r(x)$, де $c \in F$, $h(x) \in F_f(x)$ і НСД$(f,h) = 1$. До поліномів, як і до цілих чисел, можна застосувати узагальнений алгоритм Евкліда і знайти $h^{-1}\ (mod\ f) \in F_f(x)$ (деталі алгоритму Евкліда для поліномів можна знайти в [6]). Крім того, оскільки $c \in F$, то існує такий елемент $c^{-1} \in F$, що $r^{-1} = c^{-1} \cdot h^{-1} \in F_f(x)$. ■

Простий поліном $f(x)$ називається **визначальним поліномом** поля $F_f(x)$.

Теорема 21. *Нехай F скінченне поле, порядок якого p, де p – просте число, а f – простий поліном над полем F степеня n. Тоді $|F_f(x)| = p^n$.*

Доведення. Із означення поля $F_f(x)$ випливає, що множина $F(x)$ складається із поліномів, степінь яких менший $n = deg(f)$, а їх коефіцієнти належать полю F. Але таких поліномів буде p^n. ■

Наслідок 4. *Для кожного простого числа p і кожного $n \in \mathcal{N}$ існує скінченне поле, яке складається із p^n елементів.*

Приклад 3.3.1. 1. Нехай F^2 – поле лишків за модулем 2. Поліном $f(x) = x^2 + +x + 1$ є простим над полем F^2. Множина $F_f^2(x)$ є полем, яке має 2^2 елементів. Їх степені менші 2 і тому довільний елемент y із цього поля має вигляд:

$$y = b_1 x + b_0,$$

де $b_i \in F^2$, $i = 0, 1$. Таблиці Келі для поля F_2^f приймають вигляд:

+	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

*	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	$x+1$	1
$x+1$	0	$x+1$	1	x

2. Одним із важливих прикладів простих полів є скінченні кільця лишків за модулем простого числа p. Це випливає з такої теореми.

Теорема 22. *Кільце лишків Z_p за модулем p буде полем тоді і тільки тоді, коли p просте число.*

Доведення. Для доведення досить встановити існування для кожного $s \in Z_p, s \neq 0$, оберненого елемента $s' \in Z_p$.

Розглянемо елементи $s, 2s, \ldots, (p-1)s$. Всі ці елементи різні і відмінні від нуля, оскільки із $s \neq 0$ випливає $ks \neq 0$ для $k = 1, 2, \ldots, p-1$ на підставі простоти числа p. А те що вони різні випливає з того, що коли припустити $ks = ls, k < l. l, k < p$, то $(l-k)s = 0$, а це невірно. Отже, послідовність елементів $s, 2s, \ldots, (p-1)s$ збігається з послідовністю переставлених яким-небудь чином елементів $1, 2, \ldots, p-1$. Зокрема, знайдеться $s', 1 \leq s' \leq p-1$, для якого $ss' = 1$, тобто s' – обернений елемент до елемента s.■

Поле Z_p будемо позначати F_p.

Наведемо таблиці Келі поля F_5. Нехай маємо множину $A = \{0, 1, 2, 3, 4\}$, на якій визначені операції додавання і множення за модулем 5. Таблиці Келі для цих операцій матимуть вигляд:

$\oplus$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\odot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Операції додавання і множення задовольняють законам асоціативності і комутативності, а нуль і одиниця є нулем і одиницею цієї алгебри. Крім того, із таблиці множення очевидним чином випливає, що ця алгебра не має дільників нуля, тобто вона є областю цілісності, а, отже, є полем. Очевидно, що побудоване таким способом поле буде простим на підставі простоти модуля p. ♠

Розглянемо ще деякі важливі властивості поля F_q, де $q = p^k$ і p – просте число. Зазначимо, що в полі F_q існує $q-1$ ненульових елементів, які відносно операції множення утворюють абелеву групу. Будемо позначати цю групу F_q^*.

Твердження 4 *Порядок довільного елемента $a \in F_q^*$ є дільником числа q-1.*

Доведення. Нехай d – найменший степінь a, для якого $a^d = 1$. Такий степінь d дійсно існує на підставі скінченності множини F_q^*: різні степені елемента a не можуть бути всі різними і якщо $a^i = = a^j, j > i$, то $a^{j-i} = 1$.

Нехай S означає множину $\{1, a, a^2, \ldots, a^{d-1}\}$ всіх різних степенів елемента a. Тоді для довільного $b \in F_q^*$ нехай bS означає клас суміжності, який складається з елементів ba^j (наприклад, $1 \cdot S = S$). Очевидно, що два класи суміжності або збігаються або не мають спільних елементів. Дійсно, якщо $b_1a^i \in b_1S$ належить класу b_2S, тобто $b_1a^i = b_2a^j$, то для довільного b_1a^m із b_1S маємо $b_1a^m = b_1a^{i+m-1} = b_2a^{j+m-i}$. Оскільки кожний клас суміжності складається з d елементів і є розбиттям множини F_q^*, то на підставі теореми Лагранжа $d|q-1$. ■

Означення 62. *Твірним елементом скінченного поля F_q називається елемент g, порядок якого дорівнює q-1, а це еквівалентно тому, що $1, g, g^2, \ldots, g^{q-2}$ є різними елементами F_q^*.*

Теорема 23. *Довільне скінченне поле F_q має твірний елемент. Якщо g – твірний елемент F_q^*, то g^j буде твірним елементом тоді і тільки тоді, коли НСД(j,q-1) ==1. Зокрема, в F_q^* існує $\varphi(q- -1)$ різних твірних елементів.*

Доведення. Припустимо, що елемент $a \in F_q^*$ має порядок d, тобто $a^d = 1$ і жодний менший степінь a не дорівнює 1. Тоді на підставі твердження 4 число d дільник $q-1$. Оскільки a^d найменший степінь, який дорівнює 1, то всі елементи $a, a^2, \ldots, a^d = = 1$ різні. Покажемо, що елементи, порядок яких d, – це всі ті $\varphi(d)$ значень a^j, для яких НСД(j,d)=1. По-перше, так як всі d різних степенів різні, то вони є множиною всіх коренів рівняння $x^d = 1$. Отже, довільний елемент, порядок якого дорівнює d, повинен знаходитися серед степенів елемента a. Але не кожний степінь a має порядок d, оскільки коли НСД$(j,d) = d' > 1$, то елемент a^j матиме менший порядок: числа d/d' і j/d' цілі і тому $(a^j)^{d/d'} = (a^d)^{j/d'} = 1$. Навпаки, покажемо, що елемент a^j при НСД(j,d)=1, має порядок d. Дійсно, припустимо, що НСД(j,d) =1 і a^j має порядок $d'' < d$. Тоді $(a^{d''})^j = (a^{d''})^d = 1$ і, отже, $(a^{d''})^{\text{НСД}(j,d)} = a^{d''} = 1$. Але $a^{d''} \neq 1$, оскільки елемент a має порядок d. Таким чином, a^j має порядок d тоді і тільки тоді, коли НСД(j,d) =1.

Це означає, що коли маємо елемент, порядок якого d, то існує тільки $\varphi(d)$ елементів, порядок яких дорівнює d. Отже, для довільного $d|(q-1)$ існує лише дві можливості: або елементів порядку d немає, або таких елементів $\varphi(d)$.

Але кожний елемент має деякий порядок $d|(q-1)$ і або 0 або $\varphi(d)$ елементів мають порядок d. На підставі тотожності Гауса дістаємо $\varphi(q-1) = \sum_{d|(q-1)} \varphi(d) = q-1$ і права частина дорівнює кількості елементів в F_q^*. Звідси і з того, що кожний елемент має певний порядок, випливає, що для кожного $d|(q-1)$ завжди знайдеться $\varphi(d)$ (і ніколи не 0) елементів, порядок яких дорівнює d. Зокрема, існує точно $\varphi(q-1)$ елементів, порядок яких $q-1$. Як встановлено вище, якщо елемент g має порядок $q-1$, то всі інші елементи порядку $q-1$ – це елемнти g^j, для яких НСД$(j,d)=1$. ■

Наслідок 5. *Для довільного простого числа p існує таке число g, що його степені пробігають всі ненульові класи лишків за модулем p.*

Приклад 3.3.2. Нехай $p = 19$, тоді $p-1 = 18$ і всі лишки за модулем 19 степенів числа 2 породжують всі елементи мультиплікативної групи F_p^*: 2,4,8,16,13,7,14,9,18,17,15,11,3,6,12,5,10,1.

Степені числа $4 = 2^2$ не будуть породжувати всі елементи групи F_p^*, так як НСД(2,18) = 2≠ 1. Дійсно, маємо: 4,16,7,9,17,11,6,5,1,4,16,7,9,17,11,6,5,1. ♠

2.5.7. Застосування полів в криптографії

Розглянемо приклад одночасного застосування методів груп і полів в побудові криптографічної системи. Нехай, як і раніше, літери алфавіту $X = \{a, b, c, d, \ldots, x, y, z\}$ англійської мови лінійно упорядковані таким чином:

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	-
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

Побудуємо поле F_{33} за допомогою простого полінома $f(x) = x^3 + 2x^2 + 1$, що дає такі таблиці додавання і множення:

+	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	1	2	0	4	5	3	7	8	6	10	11	9	13	14	12	16	17	15	19	20	18	22	23	21	25	26	24
2	2	0	1	5	3	4	8	6	7	11	9	10	14	12	13	17	15	16	20	18	19	23	21	22	26	24	25
3	3	4	5	6	7	8	0	1	2	12	13	14	15	16	17	9	10	11	21	22	23	24	25	26	18	19	20
4	4	5	3	7	8	6	1	2	0	13	14	12	16	17	15	10	11	19	22	23	21	25	26	24	19	20	18
5	5	3	4	8	6	7	2	0	1	14	12	13	17	15	16	11	9	10	23	21	22	26	24	25	20	18	19
6	6	7	8	0	1	2	3	4	5	15	16	17	9	10	11	12	13	14	24	25	26	18	19	20	21	22	23
7	7	8	6	1	2	0	4	5	3	16	17	15	10	11	9	13	14	12	25	26	24	19	20	18	22	23	21
8	8	6	7	2	0	1	5	3	4	17	15	16	11	9	10	14	12	13	26	24	25	20	18	19	23	21	22
9	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	0	1	2	3	4	5	6	7	8
10	10	11	9	13	14	12	16	17	15	19	20	18	22	23	21	25	26	24	1	2	0	4	5	3	7	8	6
11	11	9	10	14	12	13	17	15	16	20	18	19	23	21	22	26	24	25	22	0	1	5	3	4	8	6	7
12	12	13	14	15	16	17	9	10	11	21	22	23	24	25	26	18	19	20	3	4	5	6	7	8	0	1	2
13	13	14	12	16	17	15	10	11	9	22	23	21	25	26	24	19	20	18	4	5	3	7	8	6	1	2	0
14	14	12	13	17	15	16	11	9	10	23	21	22	26	24	25	20	18	19	5	3	4	8	6	7	2	0	1
15	15	16	17	9	10	11	12	13	14	24	25	26	18	19	20	21	22	23	6	7	8	0	1	2	3	4	5
16	16	17	15	10	11	9	13	14	12	25	26	24	19	20	18	22	23	21	7	8	6	1	2	0	4	5	3
17	17	15	16	11	9	10	14	12	13	26	24	25	20	18	19	23	21	22	8	6	7	2	0	1	5	3	4
18	18	19	20	21	22	23	24	25	26	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
19	19	20	18	22	23	21	25	26	24	1	2	0	4	5	3	7	8	6	10	11	9	13	14	12	16	17	15
20	20	18	19	23	21	22	26	24	25	2	0	1	5	3	4	8	6	7	11	9	10	14	12	13	17	15	16
21	21	22	23	24	25	26	18	19	20	3	4	5	6	7	8	0	1	2	12	13	14	15	16	17	9	10	11
22	22	23	21	25	26	24	19	20	18	4	5	3	7	8	6	1	2	0	13	14	12	16	17	15	10	11	9
23	23	21	22	26	24	25	20	18	19	5	3	4	8	6	7	2	0	1	14	12	13	17	15	16	11	9	10
24	24	25	26	18	19	20	21	22	23	6	7	8	0	1	2	3	4	5	15	16	17	9	10	11	12	13	14
25	25	26	24	19	20	18	22	23	21	7	8	6	1	0	2	4	5	3	16	17	15	10	11	9	13	14	12
26	26	24	25	20	18	19	23	21	22	8	6	7	2	0	1	5	3	4	17	15	16	11	9	10	14	12	13

×	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
2	0	2	1	6	8	7	3	5	4	18	20	19	24	26	25	21	23	22	9	11	10	15	17	16	12	14	13
3	0	3	6	9	12	15	18	21	24	11	14	17	20	23	26	2	5	8	19	22	25	1	4	7	10	13	16
4	0	4	8	12	16	11	24	19	23	20	21	25	5	6	1	17	9	13	10	14	15	22	26	18	7	2	3
5	0	5	7	15	11	13	21	26	19	2	4	6	17	10	12	23	25	18	1	3	8	16	9	14	22	24	20
6	0	6	3	18	24	21	9	15	12	19	25	22	10	16	13	1	7	4	11	17	14	2	8	5	20	26	23
7	0	7	5	21	19	26	15	13	11	1	8	3	22	20	24	16	14	9	2	6	4	23	18	25	17	12	10
8	0	8	4	24	23	19	12	11	16	10	15	14	7	3	2	22	18	26	20	25	21	17	13	9	5	1	6
9	0	9	18	11	20	2	19	1	10	17	26	8	25	7	16	6	15	24	22	4	13	3	12	21	14	23	5
10	0	10	20	14	21	4	25	8	15	26	6	16	1	11	18	12	22	5	13	23	3	24	7	17	2	9	19
11	0	11	19	17	25	6	22	3	14	8	16	24	13	21	5	18	2	10	4	12	23	9	20	1	26	7	15
12	0	12	24	20	5	17	10	22	7	25	1	13	15	18	3	8	11	23	14	26	2	4	16	19	21	6	9
13	0	13	26	23	6	10	16	20	3	7	11	21	18	4	17	14	24	1	5	15	19	25	2	12	9	22	8
14	0	14	25	26	1	12	13	24	2	16	18	5	3	17	19	20	4	15	23	7	9	10	21	8	6	11	22
15	0	15	21	2	17	23	1	16	22	6	12	18	8	14	20	7	13	19	3	9	24	5	11	26	4	10	25
16	0	16	23	5	9	25	7	14	18	15	22	2	11	24	4	13	20	6	21	1	17	26	3	10	19	8	12
17	0	17	22	8	13	18	4	9	26	24	5	10	23	1	15	19	6	14	12	20	7	11	25	3	16	21	2
18	0	18	9	19	10	1	11	2	20	22	13	4	14	5	23	3	21	12	17	8	26	6	24	15	25	16	7
19	0	19	11	22	14	3	17	6	25	4	23	12	26	15	7	9	1	20	8	24	16	18	10	2	13	5	21
20	0	20	10	25	15	8	14	4	21	13	3	23	2	19	9	24	17	7	26	16	6	12	5	22	1	18	11
21	0	21	15	1	22	16	2	23	17	3	24	9	4	25	10	5	26	11	6	18	12	7	19	13	8	20	14
22	0	22	17	4	26	9	8	18	13	12	7	20	16	2	21	11	3	25	24	10	5	19	14	6	23	15	1
23	0	23	16	7	18	14	5	25	9	21	17	1	19	12	8	26	10	3	15	2	22	13	6	20	11	4	24
24	0	24	12	10	7	22	20	17	5	14	2	26	21	9	6	4	19	16	25	13	1	8	23	11	15	3	18
25	0	25	14	13	2	24	26	12	1	23	9	7	6	22	11	10	8	21	16	5	18	20	15	4	3	19	17
26	0	26	13	16	3	20	23	10	6	5	19	15	9	8	22	25	12	2	7	21	11	14	1	24	18	17	4

Остачі від ділення на $f(x)$ позначені такими числами: x – числом 3, $x+1$ – числом 4, $x+2$ – числом 5, $2x$ – числом 6, $2x+1$ – числом 7, $2x+2$ – числом 8, x^2 – числом 9, x^2+1 – числом 10, x^2+2 – числом 11, x^2+x – числом 12, x^2+x+1 – числом 13, x^2+x+2 – числом 14, x^2+2x – числом 15, x^2x+1 – числом 16, x^+2x+2 – числом 17, $2x^2$ – числом 18, $2x^2+1$ – числом 19, $2x^2+2$ – числом 20, $2x^2+x$ – числом 21, $2x^2+x+1$ – числом 22, $2x^2+x+2$ – числом 23, $2x^2+2x$ – числом 24, $2x^2+2x+1$ – числом 25, $2x^2+2x+2$ – числом 26.

Додамо до 26 літер англійського алфавіту 27-й символ "-". Нехай потрібно зашифрувати текст з ключовим словом "welcome":

"meeting in twelve"

Користуючись таблицями поля F_{3^3} дістаємо таку шифрограму:

w	e	l	c	o	m	e	w	e	l	c	o	m	e	w	e	l
m	e	e	t	i	n	g	-	i	n	-	t	w	e	l	v	e
23	4	11	4	14	12	4	23	4	11	4	14	12	4	23	4	11
12	4	4	19	8	13	6	26	8	13	26	19	22	4	11	21	4
8	8	12	23	10	25	1	10	0	21	18	3	7	8	4	25	12
i	i	m	x	k	z	b	k	a	v	s	d	h	i	l	z	m

Дешифрація відбувається таким чином: виписуємо цифри, які відповідають ключовому слову і цифри шифрограми, тобто

w	e	l	c	o	m	e	w	e	l	c	o	m	e	w	e	l
23	4	11	4	14	12	4	23	4	11	4	14	12	4	23	4	11
8	8	12	23	10	25	1	10	0	21	18	3	7	8	4	25	12
12	4	4	19	8	13	6	26	8	13	26	19	22	4	11	21	4
m	e	e	t	i	n	g	-	i	n	-	t	w	e	l	v	e

В рядку таблиці додавання, який відповідає значенню 23, знаходимо значення 8. Тоді у верхньому рядку таблиці знаходимо стовпчик, в якому знайдено значення 8. Це буде номер першої літери тексту явного (в даному випадку це буде число 12, якому відповідає літера m. Продовжуючи діяти таким чином, знаходимо явний текст.

Зашифрувати цей текст можна користуючись таблицею множення:

w	e	l	c	o	m	e	w	e	l	c	o	m	e	w	e	l
m	e	e	t	i	n	g	-	i	n	-	t	w	e	l	v	e
23	4	11	4	14	12	4	23	4	11	4	14	12	4	23	4	11
12	4	4	19	8	13	6	26	8	13	26	19	22	4	11	21	4
19	16	25	14	2	18	24	24	23	21	3	7	16	16	1	22	1
t	q	z	o	c	s	y	y	x	v	d	h	q	q	b	w	b

Одержані шифрогпами можна зашифрувати шифром Цезаря або іншим шифром, який використовує групові властивості підстановок.

Контрольні питання

1. Яке відношення називається відношенням конгруентності?
2. Дайте означення групи, абелевої групи та їх підгруп.
3. Яке найбільше число підгруп може мати група четвертого порядку?
6. Дайте означення гомоморфізму (ізоморфізму) груп, кілець, полів.
7. Яка різниця між кільцем і полем?
9. Навести приклад кільця з дільниками нуля.

Задачі і вправи

1. На множині $A = \{1, 2, 3, 4, 6, 12\}$ визначені бінарні операції

$$a * b = \text{НСД}(a, b) \text{ і } a \circ b = \text{НСК}(a, b).$$

Побудувати таблиці Келі для цих операцій.

2. На множині A задана бінарна операція * з такими властивостями: $(\forall x, y, z \in \in A)\ x * (y * z) = y * (z * x)$; а із $x * y = x * z$ випливає $y = z$. Довести, що операція * комутативна і асоціативна.

3. На множині дійсних чисел D визначена бінарна операція * з такими властивостями: $(\forall x, y, z \in D)\ x * (x * x) = x$, $x * (y * z) = x * y + z$. Довести, що $x * x = 0$, $x * 0 = x$, $x * y = x - y$.

4. а) Довести, що в групі одиничний елемент єдиний, і що довільний елемент групи має єдиний обернений елемент.

б) На множині $(R \setminus \{0\}) \times R$, де R – множина раціональних чисел, визначена операція $(a, b) * (c, d) = (ac, bc + d)$. Довести, що алгебра $G = (\{(R \setminus \{0\}) \times R\}, \{*\})$ буде групою;

в) Довести, що коли таблиця Келі є таблицею множення скінченної групи, то кожний елемент цієї групи зустрічається в кожному рядку і кожному стовпчику рівно один раз;

г) Знайти всі підгрупи групи Z_{12} лишків відносно операції додавання за модулем 12. Знайти всі твірні цієї групи. Чи є ця група циклічною?

д) Довести, що коли циклічна група має порядок n і d – дільник n, то ця група має лише одну підгрупу, порядок якої d;

е) Довести, що коли порядок циклічної групи дорівнює n, то вона має $\varphi(n)$ твірних елементів.

5. Довести, що довільна ненульова підгрупа адитивної циклічної групи сама є циклічною групою.

6. Чи може група бути ізоморфною своїй власній підгрупі? Навести приклад і відповідне обґрунтування.

7. Нехай G – група і a – її фіксований елемент. Якщо x – довільний елемент групи G, то визначимо відображення $f : G \to G$ формулою $f(x) = axa^{-1}$. Довести, що f – автоморфізм групи G.

8. Довести, що довільна скінченна група, яка складається із n елементів, ізоморфна симетричній групі підстановок деякої множини M із n елементів.

9. а) Побудувати групу підстановок, яка ізоморфна циклічній групі четвертого порядку $G = \{e, a, a^2, a^3\}$;

б) Перевірити, чи виконуються такі рівності в цій групі підстановок $f_2^2 = f_3$, $f_3^2 = = e$, $f_2^3 = f_4$, $f_2 f_4 = e$, де $f_1 = e$ – тотожна підстановка, а f_2, f_3, f_4 – решта елементів групи підстановок.

10. Довести, що скінченна напівгрупа з одиницею і законом (лівого або правого) скорочення є групою.

11. Довести, що підгрупа H групи G індексу 2 буде нормальним дільником групи G.

12. Нехай група $G = \{e, g_1, g_2, \ldots\}$ і $x \in G$ – довільний елемент групи. Довести, що множина $S = \{e, xg_1x^{-1}, xg_2x^{-1}, \ldots\}$ включає всі елементи групи G.

13. Нехай R і S – підгрупи групи G. Визначимо добуток двох підгруп R і S таким чином: $R \cdot S = \{rs | r \in R \text{ і } s \in S\}$. Довести, що

а) множина $R \cdot S$ буде підгрупою групи G тоді і тільки тоді, коли $R \cdot S = S \cdot R$;

б) якщо одна із підгруп (R чи S) є нормальним дільником групи G, то $R \cdot S = S \cdot R$ буде підгрупою групи G.

14. Показати, що порядок симетричної групи n-го степеня дорівнює $n!$.

15. Побудувати кільце для а) "трійкової" (ліва таблиця) і б) "п'ятіркової" арифметики (права таблиця)

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1,

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

16. Знайти розклад $(a+b)^3$ в некомутативному кільці.

17. Побудувати поле, порядок якого 3^2, над полем F_3 лишків за модулем 3.

18. Знайти підстановку x з рівняння $f \cdot x \cdot f_1 = f_2$, якщо

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 3 & 2 & 1 & 6 & 5 & 4 \end{pmatrix}, \quad f_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 1 & 2 & 7 & 4 & 5 & 6 \end{pmatrix},$$

$$f_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 1 & 3 & 4 & 4 & 7 & 2 \end{pmatrix}$$

Відповідь:

$$x = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 6 & 7 & 1 & 3 & 5 \end{pmatrix}.$$

19. Знайти розклад на цикли підстановок

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 4 & 5 & 1 & 6 \end{pmatrix}, \quad f_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 1 & 5 & 4 & 6 \end{pmatrix}$$

20. Довести, що коли підстановка f має розклад на цикли $s_1 s_2 \cdots s_m$, то підстановка f^k матиме розклад на цикли $s_1^k s_2^k \cdots s_m^k$.

21. Довести, що коли цикл s має довжину k, то $s^k = \varepsilon$, де ε – тотожна підстановка.

Користуючись цією властивістю, знайти f^{100}, де

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 3 & 5 & 4 & 1 & 7 & 10 & 2 & 6 & 9 & 8 \end{pmatrix}.$$

Відповідь: (див. теорему 11) $f^{100} = f$.

22. Розглянемо множину цілих чисел $\mathcal{Z}$ з такими операціями $m \oplus n = v + n - 1$ і $m \odot n = m + n - mn$, тобто алгебру $G = (\mathcal{Z}, \Omega = \{\oplus, \odot\})$.

а) Довести, що G є кільцем;

б) Знайти нульовий і одиничний елементи цього кільця;

в) Чи утворює множина непарних чисел підкільце кільця G?

г) Чи має це кільце дільники нуля?

д) Знайти ідеали цього підкільця.

23) Нехай $G = (B(U), \Omega = \{\div, \cap\})$, де $B(U)$ – булеан множини U, а $\div$ і $\cap$ – операції симетричної різниці і перетину множин відповідно.

а) Довести, що G комутативне кільце з одиницею;

б) Знайти вигляд оберненого елемента для даного елемента $A \subseteq U$;

в) Побудувати таблиці Келі для операцій цього кільця, якщо $U = \{1, 2\}$, і знайти його ідеали.

24) Нехай G – кільце з одиницею і $a, b \in G$ – дільники одиниці цього кільця. Довести, що

а) $(ab)^{-1} = b^{-1}a^{-1}$; б) одиниця кільця не є дільником нуля;

в) коли a дільник одиниці, то і $-a$ дільник одиниці.

25) Скільки ділників одиниці і нуля має кільце лишків а) Z_8; б) Z_{15}; в) Z_{23}?

26) Довести, що множина дільників одиниці кільця утворює абелеву групу відносно операції множення.

2.6. Системи обміну ключами

Познайомившись з основними алгебраїчними структурами, розглянемо алгоритми обміну інформацією між абонентами, які грунтуються на властивостях груп і полів. Отже, необхідно розв'язаты проблему обміну таємними ключами між абонентами.