Дослідження простору асоціативних пар в контексті бази знань електронного підручника

Бісікало Олег Володимирович
Вінницький державний аграрний університет

Сучасний світ вже не мислиться без Інтернет – технологій, які знаходять все більше застосування в усіх сферах людської діяльності. Тенденції розвитку спільного інформаційного простору надають особливої ***актуальності*** дослідженням в галузі розробки математичного забезпечення автоматизованих засобів передачі знань, в тому числі на основі вербальної інформації. Традиційно більшість досліджень в області розробки лінгвістичного процесору проводилося в межах моделей логічних та евристичних моделей, баз даних, експертних систем (баз знань). Найбільшою ***проблемою*** даного напрямку лишається семантичний аналіз натурально мовних конструкцій, а використання моделі образного мислення з метою покращення якісних характеристик такого аналізу має певну новизну.

Підхід до моделювання образного мислення, що пропонується [1], дозволяє формалізувати текстову інформацію бази знань на основі множини бінарних відношень між образами тексту. Фактично математична модель належить до простору впорядкованих пар, які, за змістом об'єкту дослідження отримали назву асоціативних пар [2]. Забезпечення функцій образного мислення алгебраїчними засобами вимагає розвинутого математичного апарату, тому ***постановкою задачі*** будемо вважати дослідження особливостей простору асоціативних пар з метою вибору найбільш зручної математичної інтерпретації.

В якості прикладу натурально мовної конструкції для проведення досліджень розглянемо відомі віршовані рядки класика російської літератури «Буря мглою небо кроет, вихри снежные крутя – то как зверь она завоет, то заплачет, как дитя». В таблиці 1 зображені образи цієї конструкції, з яких створюються впорядковані пари за допомогою питальних займенників [3] від головного образу (рядки таблиці) до підлеглого (стовпчики таблиці). Кожний образ таблиці має свій порядковий номер.

Таблиця 1. Наскрізний приклад асоціативних пар натурально мовної конструкції.

		буря	мглою	небо	кроет	вихри	снежные	крутя	завоет	зверь	заплачет	дитя
		1	2	3	4	5	6	7	8	9	10	11
буря	1				Что делает?				Что делает?		Что делает?	
мглою	2											
небо	3											
кроет	4	Кто?	Чем?	Что?				Как?				
вихри	5						Какие?					
снежные	6											
крутя	7					Что?						
завоет	8	Кто?								Как?		
зверь	9											
заплачет	10	Кто?										Как?
дитя	11											

Згідно з [3] до головних функцій моделі образного мислення належить накопичення з синтагматичної інформації такої конструкції, як асоціативна пам'ять та забезпечення на її основі асоціативного пошуку окремих образів. Тому математична інтерпретація повинна, з одного боку, надавати можливість об'єднувати асоціативні пари різних синтагм у складніші конструкції, а, з іншого боку, знаходити спільні пари (перетин) таких синтагм та/або конструкцій.

В якості можливих інтерпретацій універсуму асоціативних пар можна запропонувати:

1. графічну ілюстрацію у вигляді «карти зоряного неба»;
2. групу матричних операцій, що відповідають марківським процесам;
3. булеан як множина–ступінь або множина можливих підмножин множини образів.

Послідовно розглянемо переваги та недоліки всіх цих трьох варіантів.

1. ***Модель «карти зоряного неба»*** надає можливість зручної графічної інтерпретації простору асоціативних пар, якщо представляти кожну пару у вигляді координат точки на площині. В цьому випадку координати точки відповідають номерам головного M и підлеглого S образів пари, а параметри точки (розмір або інтенсивність кольору) можуть характеризувати частоту використання даної пари в асоціативній пам'яті. Розділивши приклад з таблиці 1 на дві синтагми A та B, поставимо у відповідність пар з головними членами речення, де фактично зв'язок двосторонній, вагу 2, а всім іншим парам – вагу 1. Отримані трійки значень (координата M, координата S, вага пари) для синтагм A та B представлено в таблиці 2. На основі даних цієї таблиці графічно зобразимо «карту зоряного неба» для синтагм A та B на рис.1.

Таблиця 2. Трійки значень для синтагм А та В.

Асоціативні пари А		
1	4	2
4	1	2
4	2	1
4	3	1
4	7	1
5	6	1
7	5	1

Асоціативні пари В		
1	8	2
1	10	2
8	1	2
10	1	2
8	9	1
10	11	1

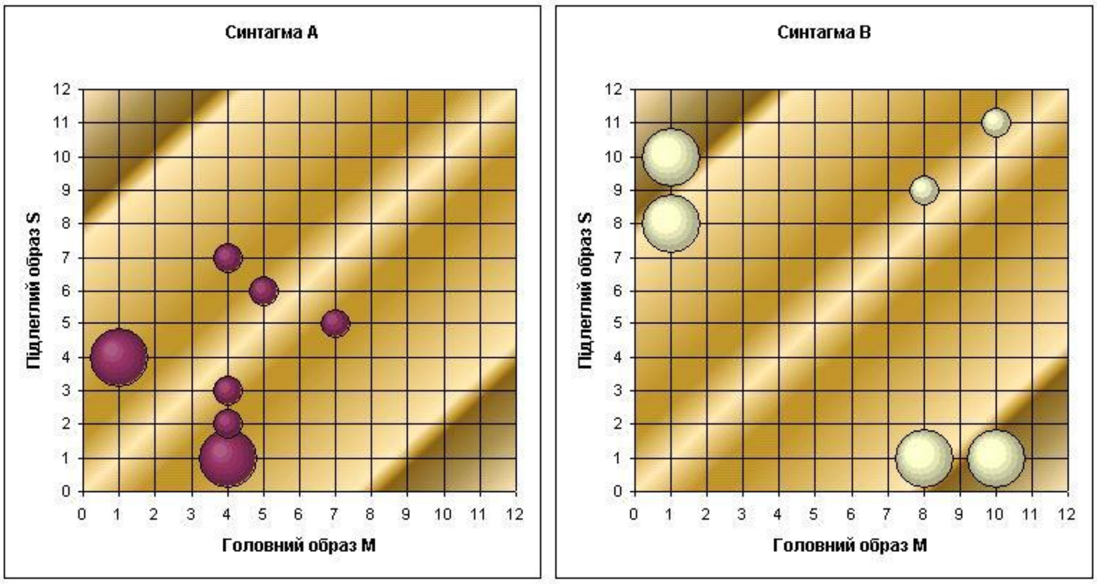

Рис.1. Зображення «карти зоряного неба» для синтагм А та В.

Визначимо основні поняття множини асоціативних пар як напівгрупи відносно операції додавання (об’єднання множин) та відносно операції множення (перетину множин). З цією метою введемо наступні визначення:

- *Операцію об’єднання* (додавання) $\cup$ будемо представляти як «накладення» двох карт; зображення «карти зоряного неба» для об’єднання синтагм А та В представлено на рис.2.

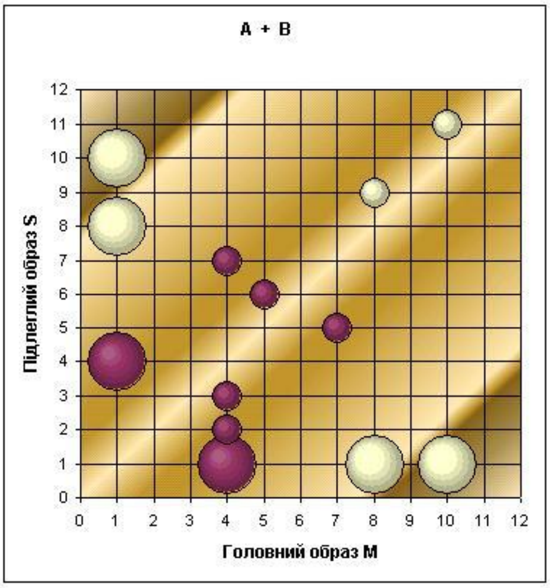

Рис.2. Зображення «карти зоряного неба» для об’єднання синтагм А та В.

- *Операцію перетину* (множення) $\cap$ будемо представляти як «порівняння» (знаходження мінімального спільного) двох карт; зрозуміло, що перетин синтагм А та В є пустою множиною.

Модель «карти зоряного неба» можна вважати напівгрупою [4] відносно операцій «накладення» та «порівняння» двох карт (фактично – об'єднання двох множин впорядкованих пар), оскільки:

1. Результат операції «накладення» або «порівняння» двох карт теж є картою

$$\forall a,b \in M \;\Big|\; a \circ b \in M\,. \qquad (1)$$

2. Операція «накладення» або «порівняння» двох карт асоціативна

$$(a \circ b) \circ c = a \circ (b \circ c)\,. \qquad (2)$$

Запропонована модель «карти зоряного неба» не може вважатися моноїдом [4], оскільки для обох визначених операцій не має сенсу нейтральний елемент *e*. Розглянемо, наприклад, *одиничний елемент 1* як симетрично «ретушовану» карту з об'єднанням прямого та зворотного зв'язку в кожній парі. Одиничний елемент може бути локальним, побудованим на основі однієї синтагми, проте для асоціативної пам'яті в цілому глобальна 1 недосяжна внаслідок динамічного характеру цієї конструкції. Граничним випадком глобальної 1 є карта, повністю заповнена максимально великими «зорями». З іншого боку, якщо *нульовий елемент 0* – симетрично «вицвіла» карта с перетином прямого та зворотного зв'язку в кожній парі, то такий нульовий елемент також може бути локальним, тобто побудованим на основі однієї синтагми. Зрозуміло, що глобальний 0 так само недосяжний для всієї асоціативній пам'яті, оскільки він вироджується в пусту карту. Локальні одиниця та нуль для синтагми А зображені на рис.3.

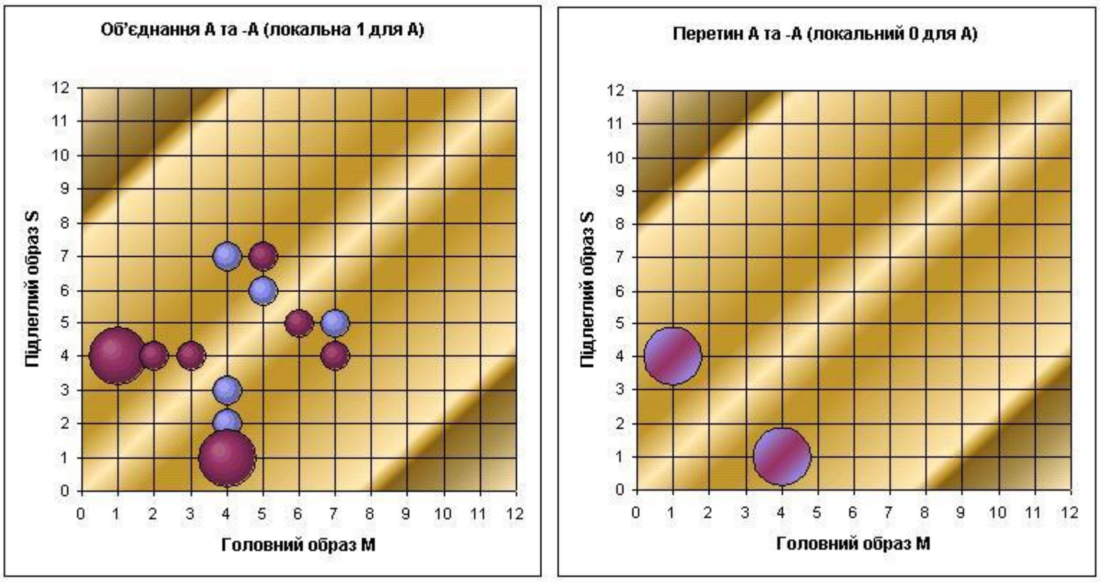

Рис.3. Зображення локальної одиниці та локального нуля для синтагми А.

При такому підході не можна вважати вірними наступні відомі співвідношення [5] для моноїдів у зв'язку з локальним характером 1 та 0:

$$x \oplus 0 = x\,; \qquad (3)$$

$$x \oplus 1 = 1\,; \qquad (4)$$

$$x \otimes 0 = 0\,; \qquad (5)$$

$$x \otimes 1 = x\,. \qquad (6)$$

Модель «карти зоряного неба» також не може вважатися групою по жодній з розглянутих операцій, оскільки по суті моделі образного мислення не має сенсу *заперечення* (зворотній елемент) у вигляді карти, симетрично перевернутій по головній діагоналі. Певний зміст можна вкласти в *заперечення* однієї асоціативної пари, коли зв'язок в ній змінює напрямок та вказує від підлеглого образу до головного. Але в результаті *заперечення* в кожній парі синтагми наступні співвідношення [5] для груп знову ж таки не виконуються у зв'язку з принциповою неможливістю досягнення глобальних 1 та 0:

$$x \oplus \bar{x} = 0\,; \qquad (7)$$

$$x \otimes \bar{x} = 1\,. \qquad (8)$$

Таким чином, модель простору асоціативних пар у вигляді «карти зоряного неба» має зрозумілу графічну інтерпретацію, проте обмежується співвідношеннями (1-2), справедливими для напівгруп.

2. Відомою моделлю дослідження випадкових процесів є ***група матричних операцій*** у вигляді однорідного марківського ланцюга, яка певною мірою схожа на попередній формалізм. У випадку моделювання асоціативної пам'яті потрібної для марківських процесів стохастичності матриць можна

досягти штучно шляхом нормування кожного рядка кінцевої матриці. В таблиці 3 представлено приклад матричних операцій з синтагмами A та B на основі такого підходу.

Таблиця 3. Матричні операції з синтагмами A та B.

Матриця A													Нормована матриця A												
	1	2	3	4	5	6	7	8	9	10	11			1	2	3	4	5	6	7	8	9	10	11	
1	0	0	0	2	0	0	0	0	0	0	0	2	1	0	0	0	1	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
4	2	1	1	0	0	0	1	0	0	0	0	5	4	0,4	0,2	0,2	0	0	0	0,2	0	0	0	0	1
5	0	0	0	0	0	1	0	0	0	0	0	1	5	0	0	0	0	0	1	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0	1	7	0	0	0	0	1	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	1	2	1	1	1	0	0	0	0	9		0,4	0,2	0,2	1	1	1	0,2	0	0	0	0	4
Матриця B													Нормована матриця B												
	1	2	3	4	5	6	7	8	9	10	11			1	2	3	4	5	6	7	8	9	10	11	
1	0	0	0	0	0	0	0	2	0	2	0	4	1	0	0	0	0	0	0	0	0,5	0	0,5	0	1
2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0
8	2	0	0	0	0	0	0	0	1	0	0	3	8	0,7	0	0	0	0	0	0	0	0,3	0	0	1
9	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0
10	2	0	0	0	0	0	0	0	0	0	1	3	10	0,7	0	0	0	0	0	0	0	0	0	0,3	1
11	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	2	1	2	1	10		1,3	0	0	0	0	0	0	0,5	0,3	0,5	0,3	3
Множення A на B													Множення нормованих матриць A та B												
	1	2	3	4	5	6	7	8	9	10	11			1	2	3	4	5	6	7	8	9	10	11	
1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	4	0	4	0	8	4	0	0	0	0	0	0	0	0,2	0	0,2	0	0,4
5	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	4	0	4	0	8		0	0	0	0	0	0	0	0,2	0	0,2	0	0,4
Додавання A та B													Нормування суми матриць A та B												
	1	2	3	4	5	6	7	8	9	10	11			1	2	3	4	5	6	7	8	9	10	11	
1	0	0	0	2	0	0	0	2	0	2	0	6	1	0	0	0	0,3	0	0	0	0,3	0	0,3	0	1
2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
4	2	1	1	0	0	0	1	0	0	0	0	5	4	0,4	0,2	0,2	0	0	0	0,2	0	0	0	0	1
5	0	0	0	0	0	1	0	0	0	0	0	1	5	0	0	0	0	0	1	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0	1	7	0	0	0	0	1	0	0	0	0	0	0	1
8	2	0	0	0	0	0	0	0	1	0	0	3	8	0,7	0	0	0	0	0	0	0	0,3	0	0	1
9	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0
10	2	0	0	0	0	0	0	0	0	0	1	3	10	0,7	0	0	0	0	0	0	0	0	0	0,3	1
11	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0
	6	1	1	2	1	1	1	2	1	2	1	19		1,7	0,2	0,2	0,3	1	1	0,2	0,3	0,3	0,3	0,3	6

Слід звернути увагу на нестохастичність матриць–синтагм для операції множення, оскільки в них, як правило, присутні пусті рядки. Проте зручність та висока ступінь дослідженості формалізму марківських процесів є перевагою цього підходу, де 1 можна вважати одиничну матрицю, 0 – нульову матрицю. Тоді матричний формалізм є моноїдом відносно операції додавання згідно з (3) та моноїдом відносно операції множення згідно з (6). Матричні перетворення також можна вважати групою відносно операції

додавання згідно з (7), якщо зворотнім елементом є матриця з від'ємними щодо початкової матриці елементами. Аналогічно матричні перетворення будемо вважати групою відносно операції множення згідно з (8), якщо зворотнім елементом є транспонована матриця.
Слід відмітити, що суттєвим недоліком даного формалізму є фізіологічна невідповідність ймовірності переходу із стану в стан марківського ланцюга процесам вибору асоціації та переходу до нового образу.
3. Розглянемо ***застосування булеану*** для формальної інтерпретації простору асоціативних пар. Будемо позначати через P(X) множину-ступінь (булеан) множини X. За визначенням [7] решітка P(X) є скороченим позначенням для

$$\{ Y \,|\, Y \subseteq X \}. \tag{9}$$

Справедливо для кінцевої множини образів X, що містить в собі n елементів, P(X) має 2^n елементів (звідси, власне, назва – множина-ступінь). Пропонується зобразити булеан на площині у вигляді графу, який ілюструє відношення часткового порядку решітки та складається з n+1 шару. На першому шарі знаходиться єдиний вузол (позначається кодом 000…000), який будемо вважати глобальним нулем. На другому шарі завжди знаходяться n вузлів з кодами 100…000, 010…000, 001…000, …, 000…010, 000…001 – можливі образи, з яких складається простір асоціативних пар. На третьому шарі також завжди маємо множину всіх можливих асоціативних пар, які можна скласти з n образів – позначення кожного вузла має дві одиниці, а в усіх інших розрядах коду знаходяться нулі. За такою самою схемою на i-му шарі кількість одиниць в коді кожного вузла дорівнює i-1, а кількість таких вузлів дорівнює числу сполучень з n по i-1:

$$C_n^{i-1} = \frac{n!}{(i-1)!(n+1-i)!}. \tag{10}$$

За логікою моделі образного мислення вузли третього шару та вище до n-го є синтагмами та/або цілими текстами. Зрозуміло, що на n+1 шарі знаходиться один вузол з одиницями в усіх n розрядах коду, який будемо вважати глобальною одиницею. Запропонована графічна інтерпретація булеану проілюстрована на рис.4 для випадків тривимірної та чотирьохвимірної решітки.

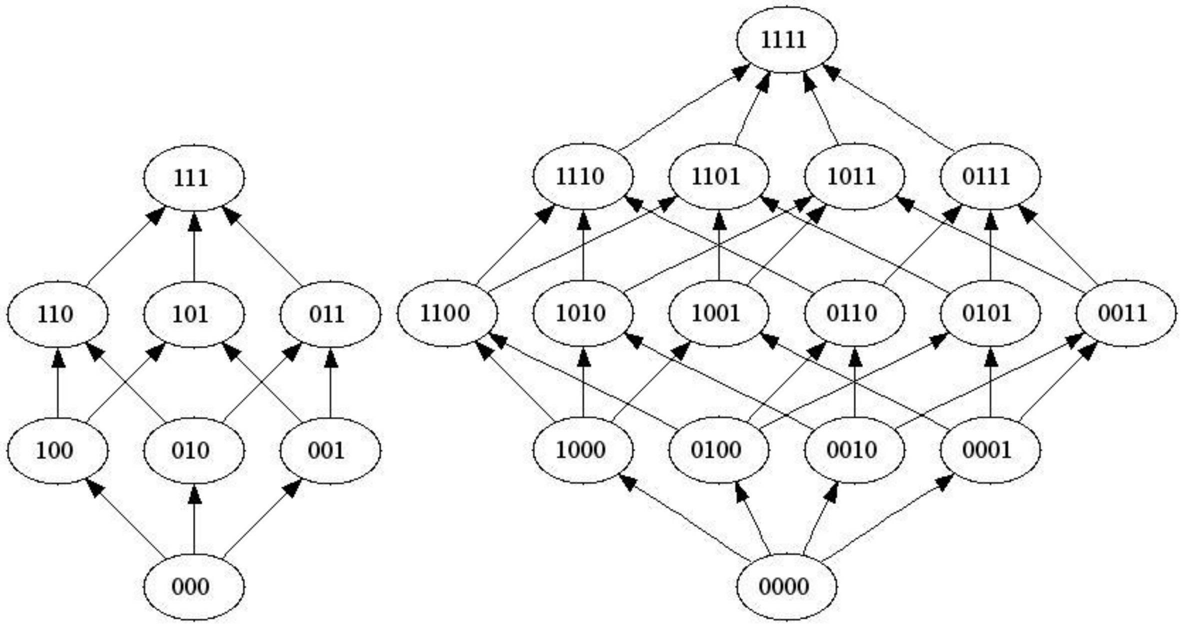

Рис.4. Багатошарова графічна проекція решітки булеану на площину для випадків n=3 та n=4.

Розглянемо на кодах вузлів булеану унарну операцію доповнення вузла, а також такі дві бінарні операції, як об'єднання двох вузлів або знаходження їх найближчого спільного нащадка та перетин двох вузлів або находження їх найближчого спільного предка. Формально ці операції задамо у вигляді таблиці істинності (табл.4) для кожного з розрядів коду.

Таблиця 4. Таблиці істинності для порозрядних операцій доповнення, об'єднання та перетину.

	Доповнення ¬	Об'єднання ∪		Перетин ∩	
Аргументи		1	0	1	0
1	0	1	1	1	0
0	1	1	0	0	0

Необхідно відмітити, що, на відміну від бінарних операцій, доповнення або операція знаходження діагонального вузла решітки булеану не має певного сенсу в моделі асоціативної пам'яті.

Формальну інтерпретацію простору асоціативних пар на основі булеану можна вважати [5] дистрибутивною обмеженою решіткою з доповненням до кожного елементу, оскільки справедливо:

a. за визначенням решітки (властивості ідемпотентності, комутативності, асоціативності та поглинання)

$$a \cup a = a, \quad a \cap a = a; \tag{11}$$

$$a \cup b = b \cup a, \quad a \cap b = b \cap a; \tag{12}$$

$$a \cup (b \cup c) = (a \cup b) \cup c, \quad a \cap (b \cap c) = (a \cap b) \cap c; \tag{13}$$

$$(a \cap b) \cup a = a, \quad (a \cup b) \cap a = a; \tag{14}$$

b. за дистрибутивністю

$$a \cup (b \cap c) = (a \cup b) \cap (a \cup c), \quad a \cap (b \cup c) = (a \cap b) \cup (a \cap c); \tag{15}$$

c. згідно з властивістю обмеженості

$$a \cup 1 = 1, \quad a \cap 0 = 0; \tag{16}$$

$$a \cup 0 = a, \quad a \cap 1 = a; \tag{17}$$

d. згідно з існуванням та властивостями доповнення

$$a \cup \neg a = 1, \quad a \cap \neg a = 0; \tag{18}$$

$$\neg(\neg a) = a; \tag{19}$$

$$\neg(a \cap b) = \neg a \cup \neg b, \quad \neg(a \cup b) = \neg a \cap \neg b. \tag{20}$$

Вираз (20) фактично представляє собою закони де Моргана, які можуть бути застосованими до простору асоціативних пар в розглянутій інтерпретації.

Висновки

В результаті аналізу трьох формальних підходів до моделювання асоціативної пам’яті визначено переваги та недоліки кожного з них. Визначено, що модель «карти зоряного неба» є напівгрупою відносно операцій об’єднання та перетину «карт», а матрична інтерпретація моделі може вважатися групою відносно операцій додавання та множення матриць. Графічна інтерпретація простору асоціативних пар у вигляді булеану представляє собою дистрибутивну обмежену решітку з доповненням до кожного елементу.

З точки зору моделювання особливостей образного мислення людини та повноти математичного апарату найбільш зручним з трьох формальних підходів можна вважати інтерпретацію простору асоціативних пар у вигляді булеану.

Перспективи

Подальші дослідження у даному напрямку варто проводити шляхом формального визначення основних понять концептуальної моделі образного мислення людини в рамках обраної математичної інтерпретації, введення і обґрунтування базових операцій, відношень та предикатів багатоосновної алгебраїчної системи.

Література

1. Бісікало О.В. Проектування архітектури бази знань експертної системи для створення електронних посібників. В збірнику «Інформаційні технології в освіті, науці і техніці / Матеріали V Всеукраїнської конференції молодих науковців ІТОНТ-2006: Черкаси, 3-5 травня 2006 р.». – Черкаси: ЧНУ, 2006. – с.98.
2. Бісікало О.В. Архітектура електронного підручника на основі бази знань навчальної експертної системи. В збірнику «Інформаційні технології в освіті: матеріали Всеукраїнської науково-практичної конференції (24-26 травня 2006 р.)». – Мелітополь, МДПУ, 2006. – с. 7-9.
3. Бісікало О.В. Принципи побудови бази знань експертної системи в галузі приладобудування // Вісник Черкаського державного технологічного університету. – Черкаси, 2006. – СПЕЦВИПУСК-2006. – с. 12-14.
4. Борисенко О.А. Лекції з дискретної математики (множини і логіка). – Суми: Університетська книга, 2002. – 180 с.
5. Новиков Ф.А. Дискретная математика для программистов. – СПб: Питер, 2001. – 304 с.: ил.
6. Капітонова Ю.В., Кривий С.Л., Летичевський О.А., Луцький Г.М., Печорін М.К. Основи дискретної математики. – Київ: Наукова думка, 2002. – 579 с.
7. Столл Роберт. Множества. Логика. Аксиоматические теории: Пер с англ. – М.: Просвещение, 1968. – 231 с.: ил.