

ІНТЕЛЕКТУАЛЬНА ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ОТРИМАННЯ ДОВІРЧИХ РІШЕНЬ ЗА ЛЮДИНОЦЕНТРОВАНИМ ПІДХОДОМ

Едуард Манзюк, Олександр Бармак, 2022 рік

1. Barmak O., Manziuk E., Krak I. Classification based hierarchical clustering prediction variability in the ensembles of models using a statistical Approach. *The 2020 IEEE 15th International Scientific and Technical Conference “Computer Science and Information Technologies”*: Proceedings (Lviv–Zbarazh (Ukraine), September 23–26, 2020). Vol. 1. Pp. 11–14. URL: <https://doi.org/10.1109/CSIT49958.2020.9322019>
2. Barmak O., Krak I., Manziuk E., Lytvynenko V., Kalyta O. Classification technology based on hyperplanes for visual analytics with implementations for different subject areas. *CEUR-WS*. 2020. Vol. 2623. Pp. 96–106.
3. Barmak O., Manziuk E., Krak I. Using piecewise hyper linear classification in multidimensional feature space for text content. *The 2019 IEEE 14th International Scientific and Technical Conference “Computer Sciences and Information Technologies”*: Proceedings (Lviv (Ukraine), September 17–20, 2019). Pp. 119–123. URL: <https://doi.org/10.1109/STC-CSIT.2019.8929798>
4. Krak I., Barmak O., Manziuk E. Using visual analytics to develop human and machine-centric models: A review of approaches and proposed information technology. *Computational Intelligence*, 2022. Vol. 38, No 3. Pp 921–946. URL: <https://doi.org/10.1111/coin.12289>.

Актуальність теми

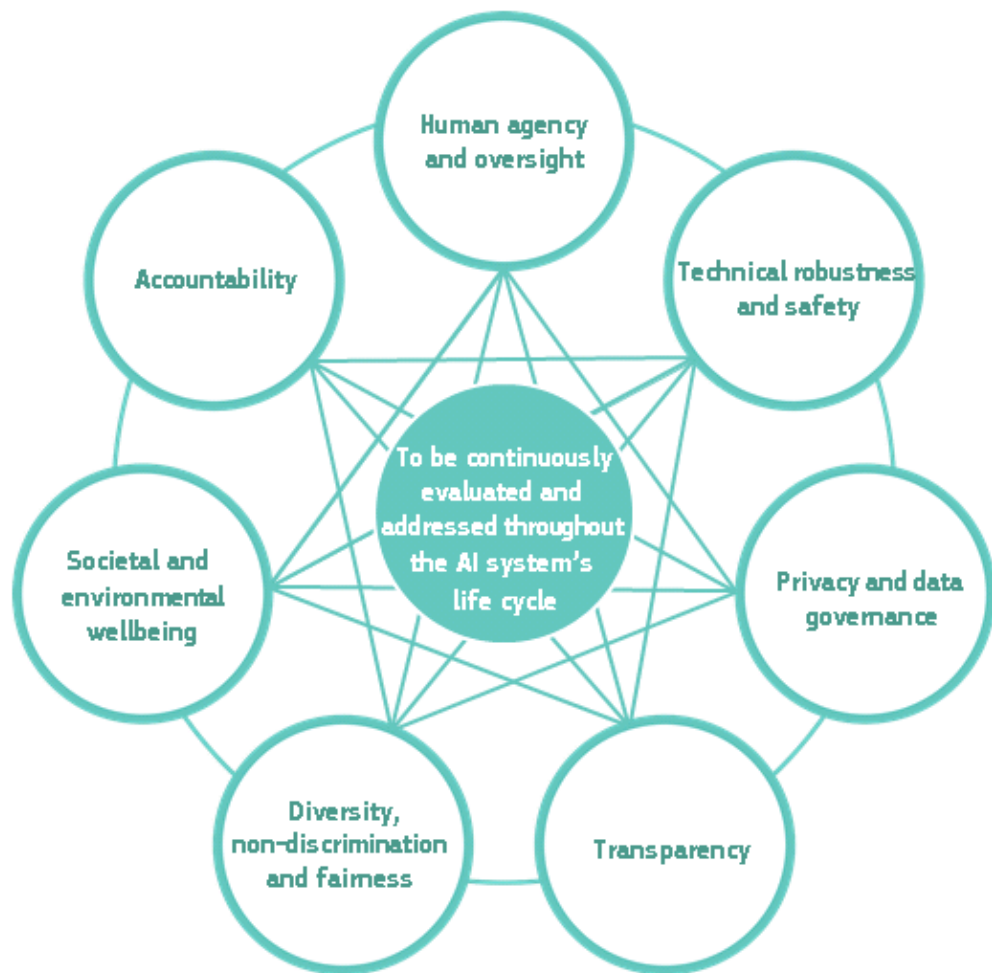
ЛЮДИНОЦЕНТРИЗМ → ЕТИЧНІ ПРИНЦИПИ → ДОВІРА ДО ІІТ

- Широке швидке поширення та застосування інтелектуальних інформаційних технологій (ІІТ).
- Впровадження система прийняття рішень на основі ІІТ.
- Делегування прийняття рішень системам ІІТ.
- Системи ІІТ набувають форм суб'єктності при прийнятті рішень.
- Використання в системах критичного значення для суспільства та людського життя.

ІІТ вимагає розробки формалізованих підходів та побудови основних принципів функціонування предметних областей використання ІІТ.

Ця потреба втілюється в розробці рекомендацій і стандартів для отримання максимальної вигоди від використання ІІТ та мінімізації можливих ризиків. Нормативна база будується на **людиноцентричній основі**.

Етичні рекомендації довіри до ІІТ

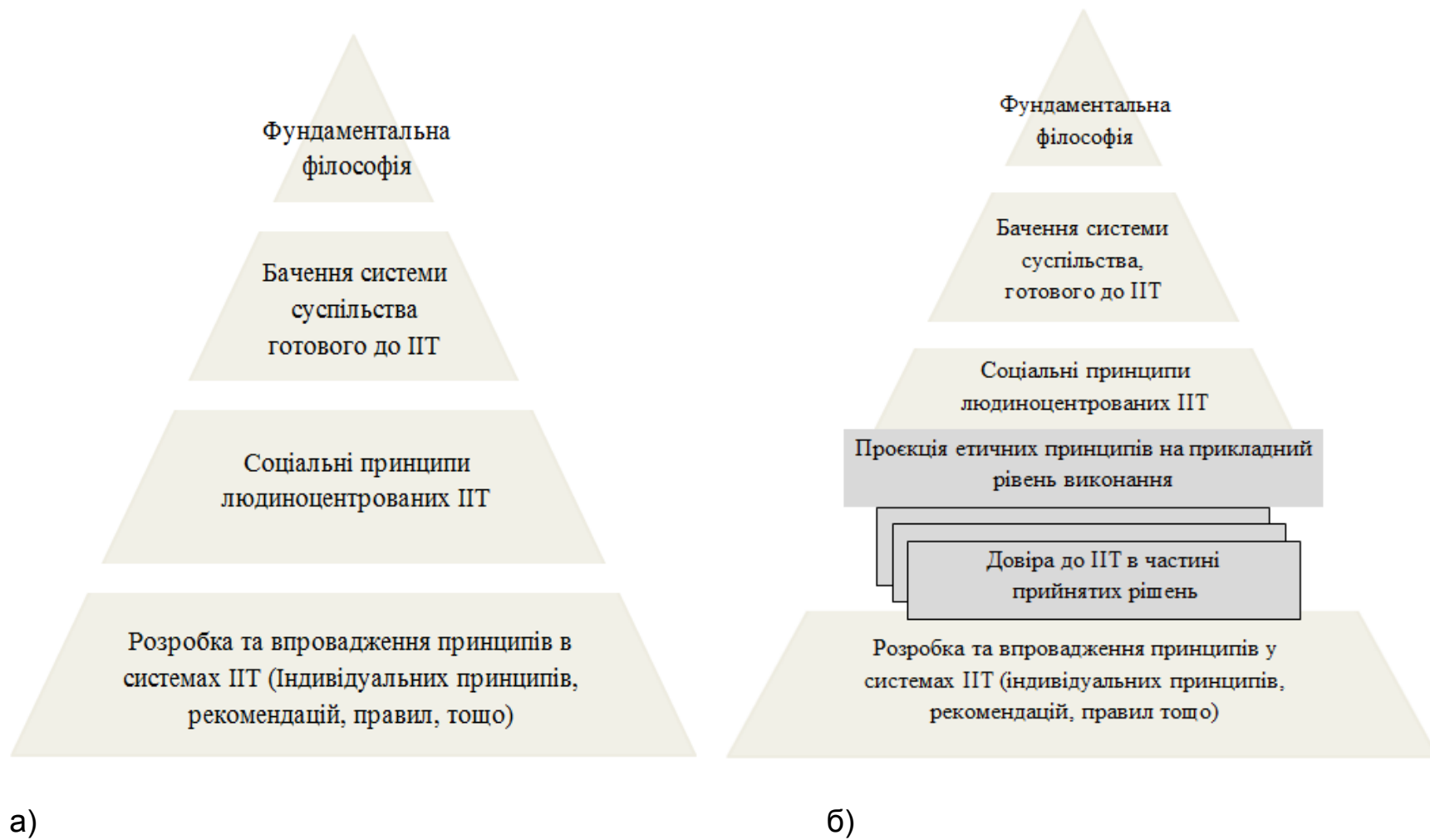


1. Розширення можливостей людини.
2. Технічна надійність та безпека.
3. Конфіденційність та управління даними.
4. Зрозумілість (прозорість).
5. Різноманітність, відсутність дискримінації та справедливність.
6. Благополуччя суспільства та навколишнього середовища.
7. Підзвітність.

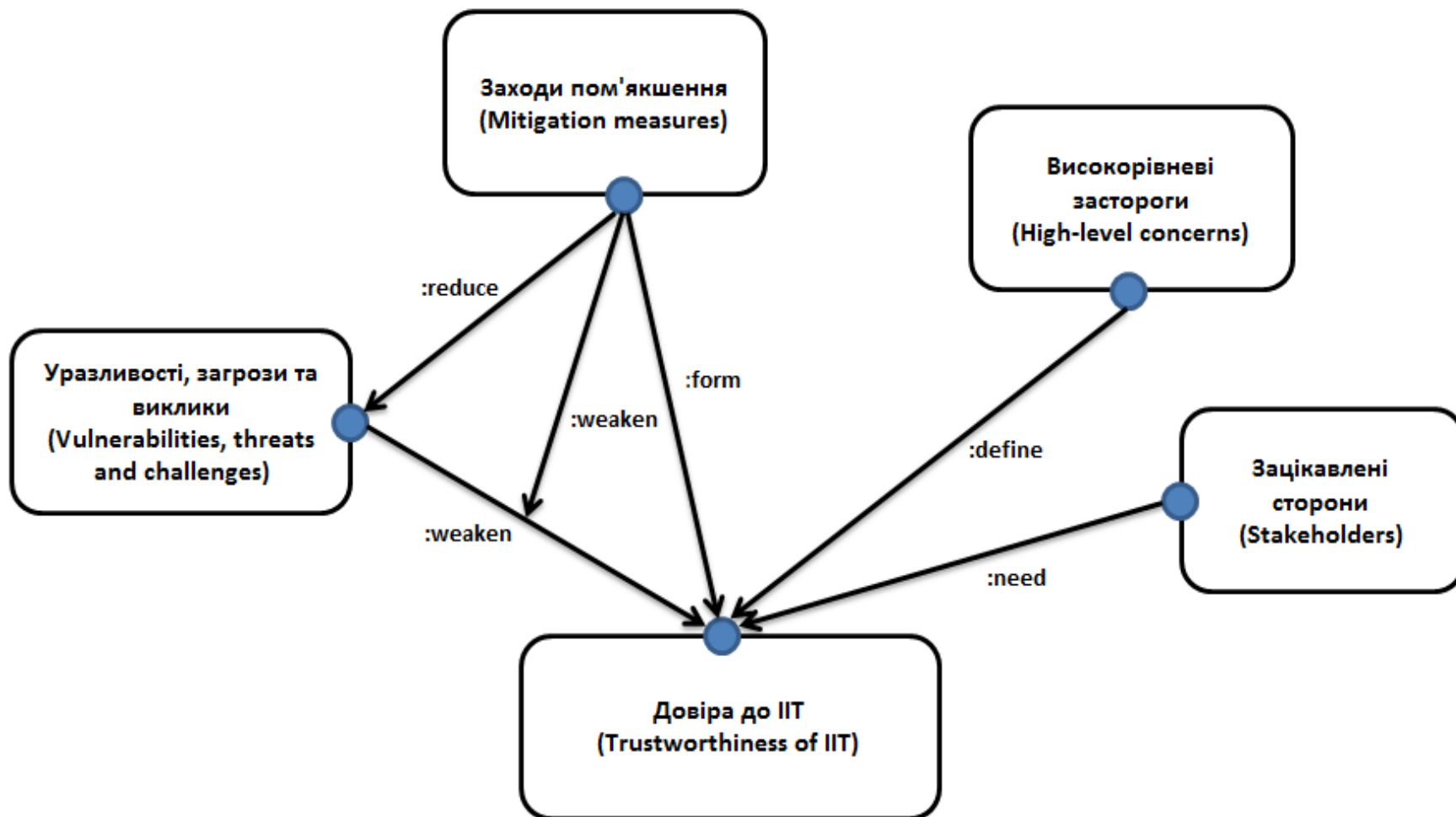
Взаємозв'язок семи вимог: всі мають однакову важливість, підтримують один одного і повинні впроваджуватися та оцінюватися протягом усього життєвого циклу систем ІІТ.

* HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE SET UP BY THE EUROPEAN COMMISSION

Довіра до ІТ як основа соціотехнологічного розвитку суспільства



Онтологія довіри до ІІТ



Онтологія довіри до ІІТ стандарту ISO/IEC TR 24028:2020

Онтологія довіри до ІІТ стандарту ISO/IEC TR 24028:2020

$$SbDmTrw = \langle Con, Rel, Int \rangle$$

Con - сукупність понять, визначених у рамках стандарту;

Rel - сукупність зв'язків між поняттями;

Int - множина функцій інтерпретації.

$$Con = \{\{Vul, Thr, Chal\}, \{MtMs\}, \{HLConc\}, \{Sth\}, \{TrW\}\}$$

$$Rel = \{"reduce", "weaken", "form", "define", "need"\}$$

Концепт «Заходи пом'якшення» *MtMs* формує довіру до ІІТ

$$TrW = f(MtMs)$$

$$MtMs = \{Ct, subCt, Ch\}$$

Ch - характеристики, *subCt*- підкатегорії, *Ct* - категорії.

$$M_{MtMs} = \left\{ \begin{array}{l} Trs, Exp, Cont, Bsred, Prv, Rlb, Rsl, Rbs, Flmit, \\ Sffun, Tst, Evl, Us, Apl, "isPartOf" \end{array} \right\}$$

зрозумілість, пояснювальність, контрольованість, стратегії зменшення упередженості, конфіденційність, безпечність, резильєнтність, робастність, пом'якшення впливу несправностей апаратного забезпечення системи, функціональна безпека, тестування, оцінювання, використання, застосовність

**Рівень синтаксичного встановлення відповідності онтології
та структурованого домену**

$$\text{syntactic}(ent, ent') = \frac{w_{label} \sigma(\text{matchClass}(ent, ent')) + \frac{\sum w_{prop_vector}(\text{matchProp}(ent, ent'))}{|prop_vector|}}{w_{label} + \frac{\sum w_{prop_vector}}{|prop_vector|}}$$

$$s.t. \text{ syntactic}(ent, ent') \neq \text{syntactic}(ent', ent);$$

$$ent \in Ent_O, ent' \in Ent_{Ds}, \text{syntactic}(ent, ent') \in [0, \dots, 1].$$

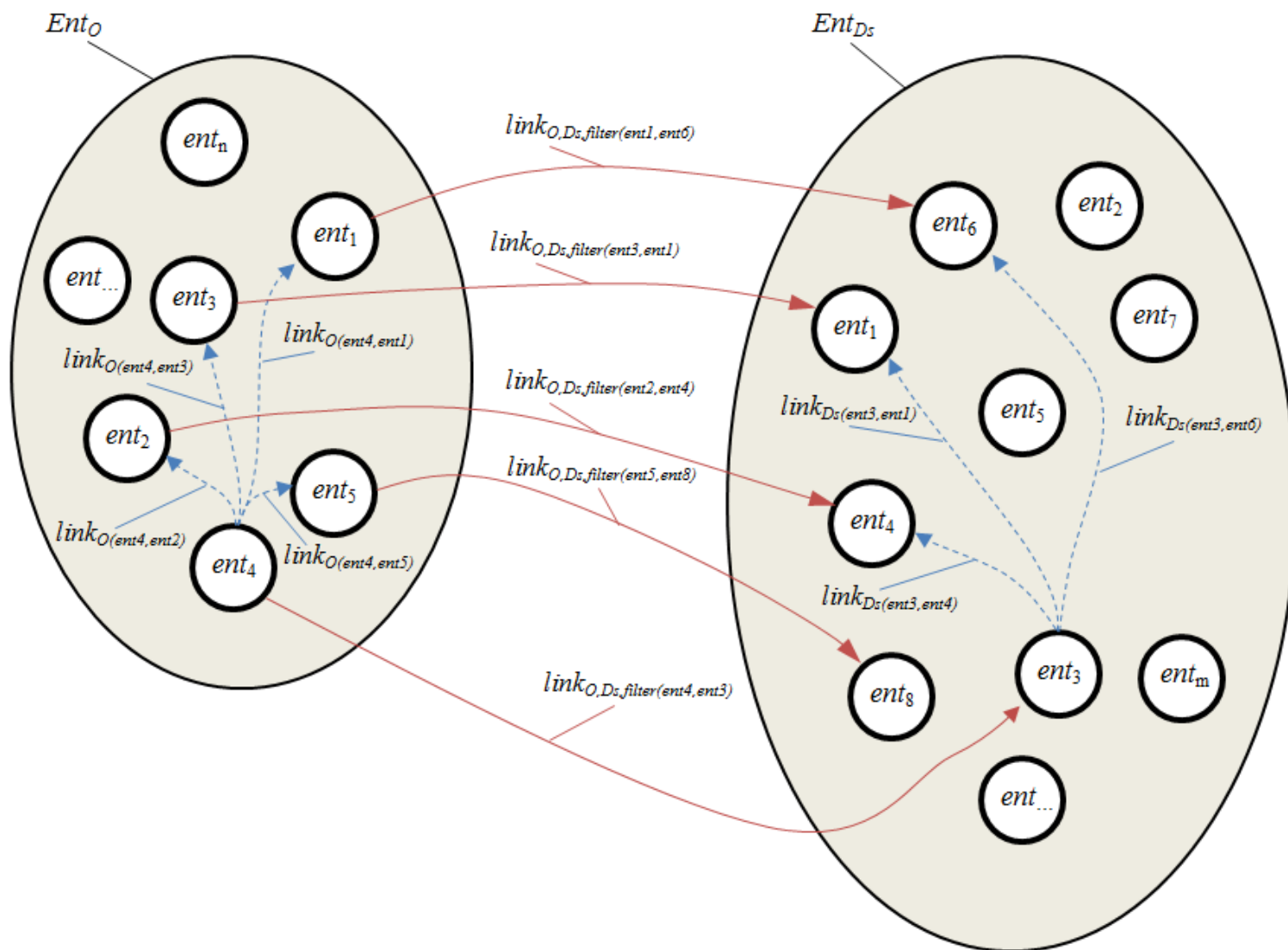
**Рівень структурного встановлення відповідності онтології та
структурованого домену**

$$\text{structure}(ent, ent') = \begin{cases} \frac{|A| + |B|}{2|\pi_{ent}(Link_O)|}; \\ A = \sigma_{link_{Ds}(ent', y) : \exists link_{O, Ds, filter(x, y)} \wedge \exists link_{O(ent, x)} \wedge (\forall x \in Ent_O) \wedge (\forall y \in Ent_{Ds})} (\pi_{ent'}(Link_{Ds})); \\ B = \sigma_{link_{O, Ds, filter(x, y)} : \exists link_{O(ent, x)} \wedge (\forall y \in Ent_{Ds}) \wedge (\forall z \in Ent_O)} (Link_{O, Ds, filter}), \end{cases}$$

$$s.t. \text{ structure}(ent, ent') \neq \text{structure}(ent', ent), \text{structure}(ent, ent') \in [0, \dots, 1],$$

$$\exists link_{O, Ds, filter(ent, ent')}, ent \in Ent_O, ent' \in Ent_{Ds}, \exists ! link = \sigma_{\max_w}(Link), w \in [0, \dots, 1].$$

Схема зв'язків структурного рівня встановлення відповідності сутностей



Рівень семантичного встановлення відповідності онтології та структурованого домену

Інформативні рамки

$$ScIn = \langle Ent, Prop(Des), Scp \rangle$$

Ent - набір іменованих сутностей; *Prop(Des)* - набір властивостей, які утворюють іменовані сутності та визначаються набором описів; *Scp* - рамки властивостей визначаються кортежем

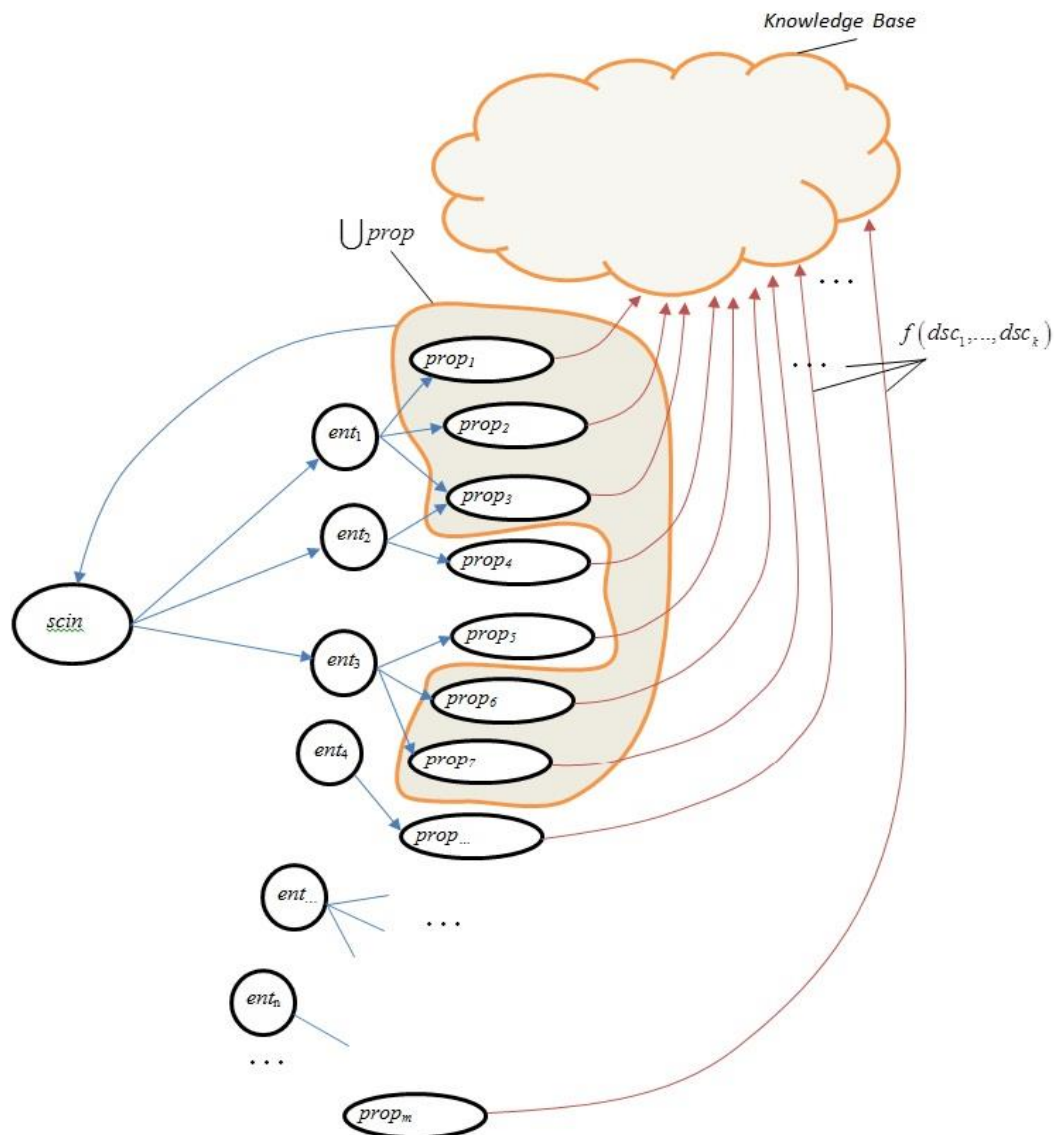
$$Scp = \{ \langle ent, Scp(ent) \rangle \mid \forall ent \in Ent, ent \text{ is entity } ScIn \}$$

$$Scp(ent) = \{ \forall prop \in Prop \mid prop \text{ is property } ent \}$$

$$semantic(ent, ent' \Rightarrow scin) = \frac{\left| \sigma_{prop_{Ds} \in Prop_{Ds} \wedge (Fun_{semanticPropertie}(prop_O, prop_{Ds}) > threshold)} (\pi_{ent}(Prop_O)) \right|}{\left| \pi_{ent}(Prop_O) \right|}$$

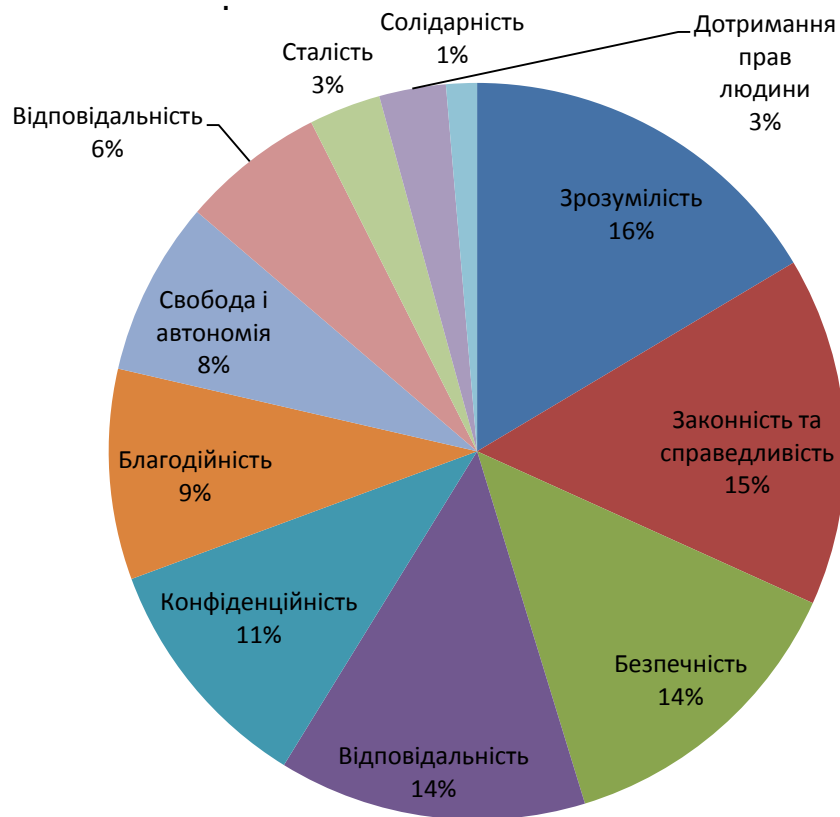
$$semanticPropertie(prop_O, prop_{Ds}) = f(\{dsc_O\}, \{dsc_{Ds}\})$$

Схема побудови елемента інформативної рамки (метасутності) у формі сукупностей властивостей як функції описів отриманих з бази знань



Експериментальні дослідження

Розподіл етичних принципів ІІТ



Згідно документів у яких визначено принцип на основі корпусу сірої літератури та soft-law документів

Проекція складових концепту довіри в площину інтелектуальних інформаційних систем як підмножини ІІТ

Основні способи забезпечення довіри до систем ІІТ на трьох рівнях архітектури:

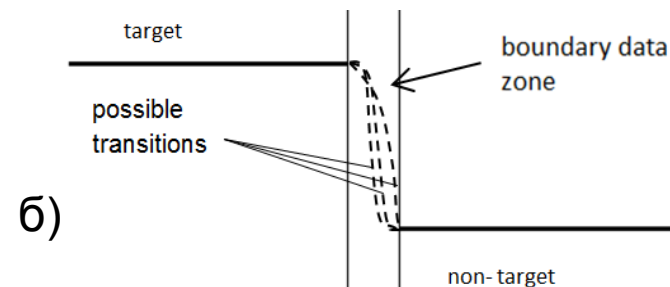
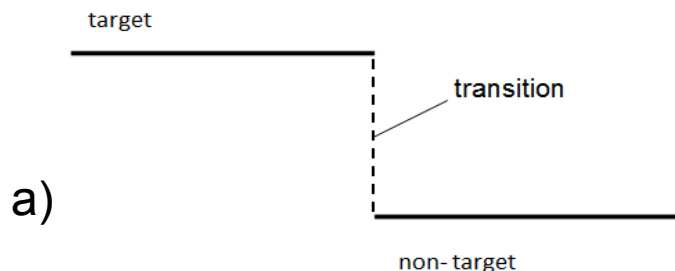
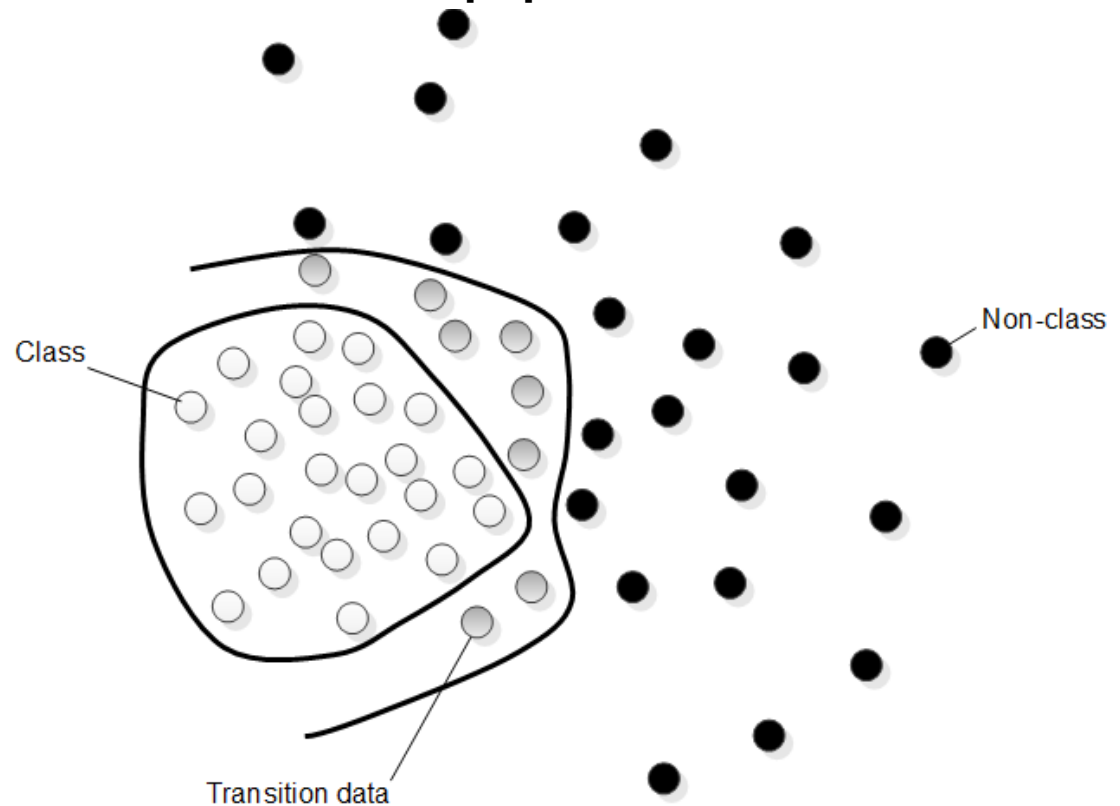
- фізичні: рівень датчиків і виконавчих механізмів;
- інфраструктурні: рівень, що включає апаратне зберігання, обробку та передачу інформації у вигляді даних;
- **прикладний**: рівень прикладного програмного забезпечення, яке реалізує алгоритми інтелектуальної обробки даних.

Складові концепту довіри на **прикладному** рівні інтелектуальної інформаційної системи:

- **інтерпретованість** (властивість внутрішнього стану системи бути зрозумілим і поясненим), **пояснювальність** та **транспарентність**, **простежуваність**;
- **всесторонність розгляду** (для забезпечення різноманітності, об'єктивності та справедливості);
- **аналізованість** (можливість встановлення рівня впливу параметрів на кінцевий результат).

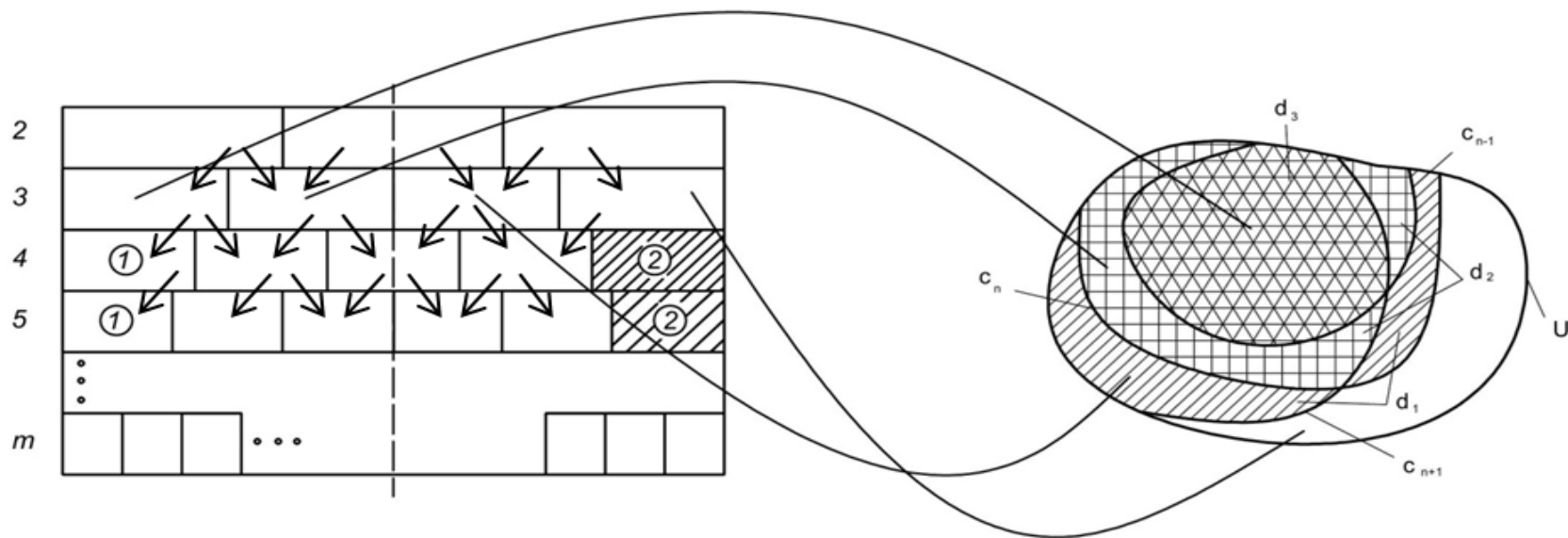
Сегментування даних на основі рішень ансамблів моделей та ієрархії групуваності і визначення інтерпретованих даних

Приклад групування в однокласовій класифікації та наявність сегменту складнокласифікованих даних



Перехід між даними: а) ідеальний перехід між класами; б) можливі форми переходу в області граничних даних.

Міграція даних при формуванні набору рішення класифікатора комбінації

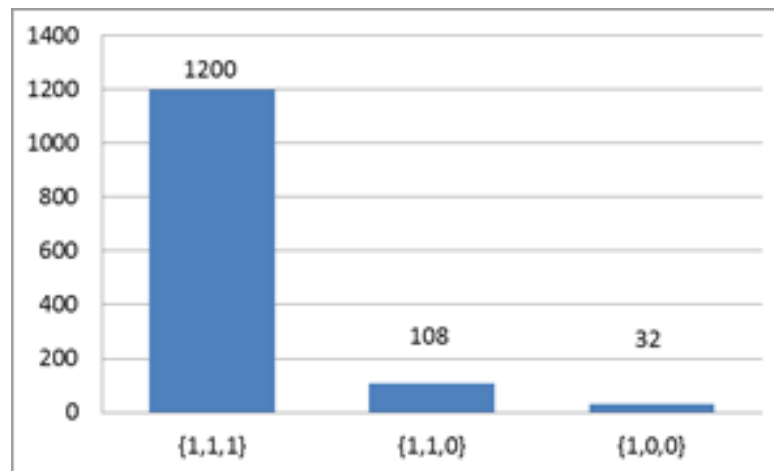


3 класифікатори $\{1,1,1\}$ $\{1,1,0\}$ $\{1,0,0\}$ $\{0,0,0\}$

d_3 d_2 d_1 d_0

перехідна зона 11,67%

$$C = \{c | A(c)\}, \quad A(c_i) = \begin{cases} c_i = \arg \max_{j \in [1, \dots, |W|]} b_j, b_j = \begin{cases} \bigcap_{k=1}^i C_k; \\ \bigcup_{k=1}^i C_k \setminus U. \end{cases} \\ \bigcup_C C \setminus U \neq \{ \} \end{cases}$$



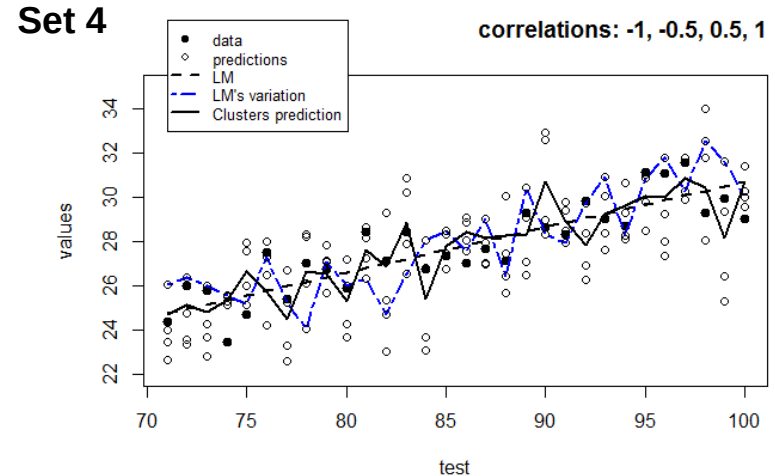
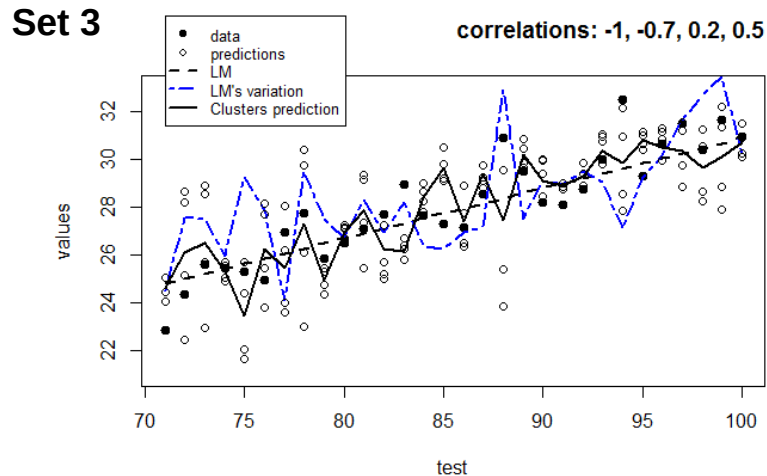
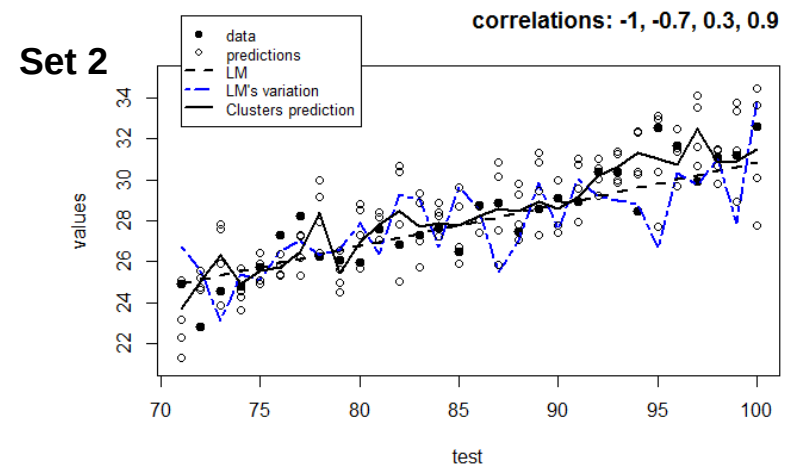
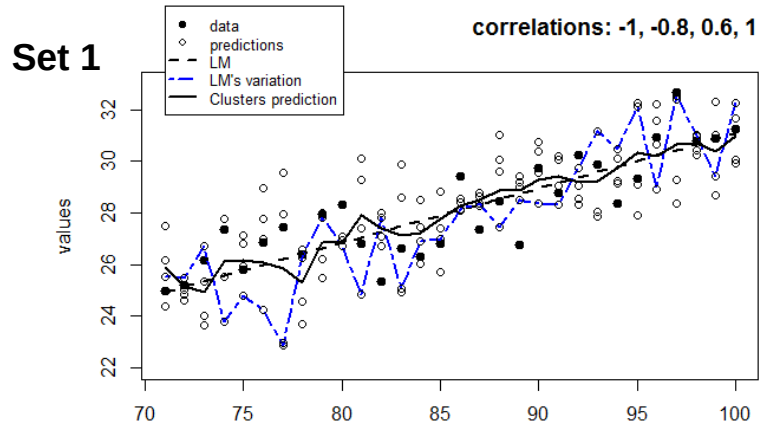
Вимоги до формування ансамблю СПР

Формування ансамблю СПР:

- СПР забезпечують високі якісні показники, для формування множини консолідованих рішень;
- СПР забезпечує наявність альтернативних рішень (за певною множиною даних), для забезпечення відсутності повного дублювання за усіма рішеннями;
- усі елементи даних повинні бути правильно класифіковані хоча б однією СПР.

Визначення ефективності та параметрів впливу щодо підходу з використання ієрархії узагальнення групових рішень на основі імітації рішень СПР ансамблів

Розподіли передбачень на основі лінійної регресії з моделюванням помилок та ієрархічним групуванням ансамблем СПР на заданих наборах кореляції.



Set 1

	bias	variance
LM	-0.1009	1.1591
LM's variation	-0.3581	2.9016
Clusters prediction	-0.0734	1.2558

Set 2

	bias	variance
LM	-0.2016	1.1859
LM's variation	-0.2024	3.7683
Clusters prediction	0.2444	1.5576

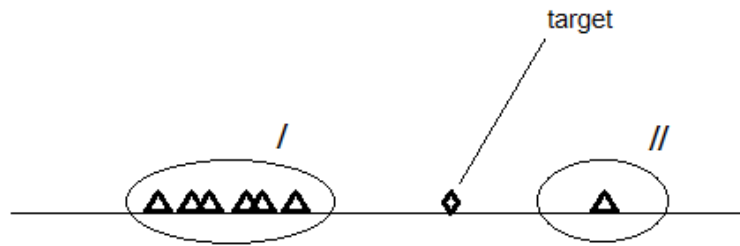
Set 3

	bias	variance
LM	-0.23351	1.0853
LM's variation	0.33731	3.7138
Clusters prediction	-0.09598	2.1010

Set 4

	bias	variance
LM	-0.03141	1.0223
LM's variation	0.20085	1.9691
Clusters prediction	-0.04302	1.4182

Співвідношення рішень СПР і зміщення рішень ансамблю



Ієрархія другого рівня і незміщена оцінка найвищого рівня.

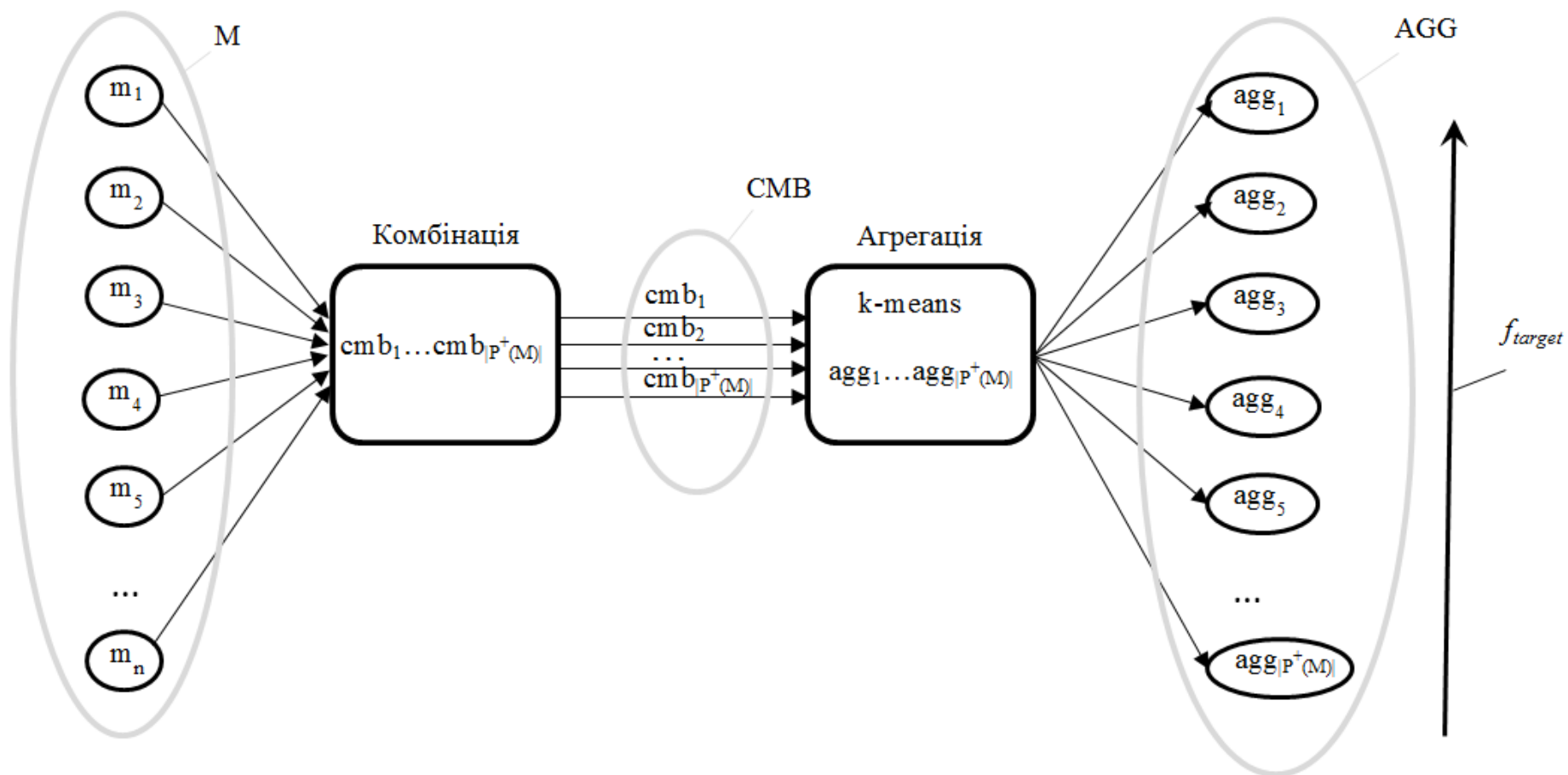
Відношення між прогнозами двох кластерів групи $C_{L1} = \{C_1, C_2, \dots, C_{n1}\}$ і $C_{L2} = \{C_1, C_2, \dots, C_{n2}\}$, $n1, n2 \in N$. Нехай $Err(C_{L1})$ і $Err(C_{L2})$ помилки передбачення двох груп моделей.

$$signP(C_{Li}(x)) \neq signP(C_{Lk}(x)), \quad i \neq k \in \{1, 2\}, \quad \eta \in [0, 1].$$

Головною умовою мінімізації похибки ансамблевого передбачення є рівність дисперсії груп ансамблю моделей і протилежність їх прогнозів.

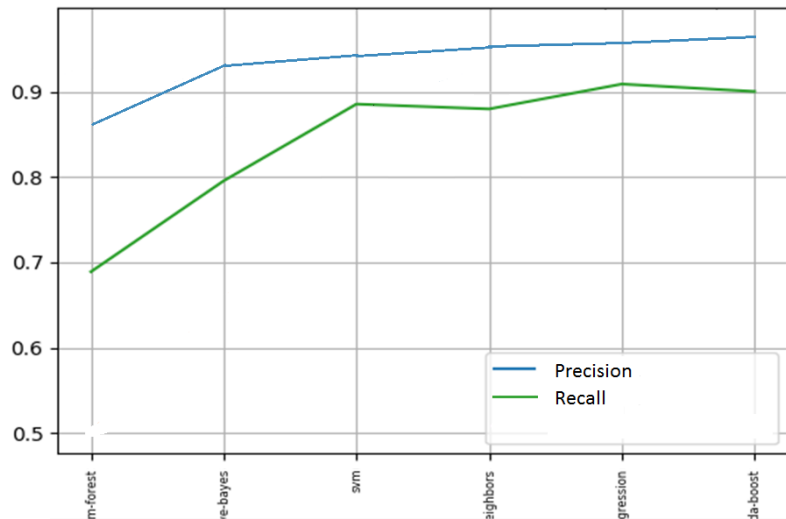
$$\begin{cases} \sigma_i^2 = \sigma_k^2, i \neq k \in \{1, 2\}; \\ signP(C_{Li}(x)) \neq signP(C_{Lk}(x)), x \in X, i \neq k \in \{1, 2\}. \end{cases}$$

Схема метод групового прийняття рішень ансамблем СПР на основі ієрархічної груповності рішень ансамблю СПР

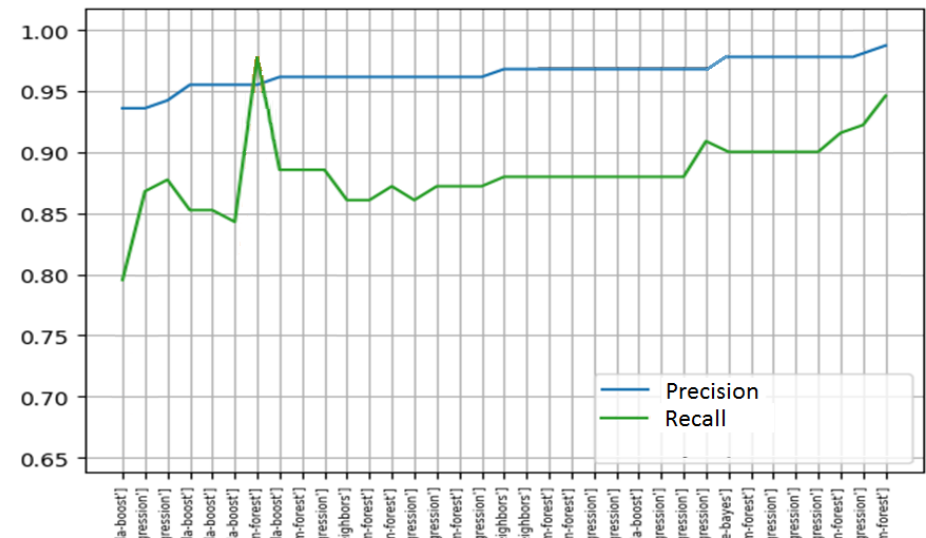


Експериментальні дослідження ефективності застосування інтелектуальної інформаційної технології отримання довірчих рішень за показниками точності та повноти (корпус Reuters-21578)

{'svm','naive-bayes','k-neighbours','ada-boost','random-forest','logistic-regression'}

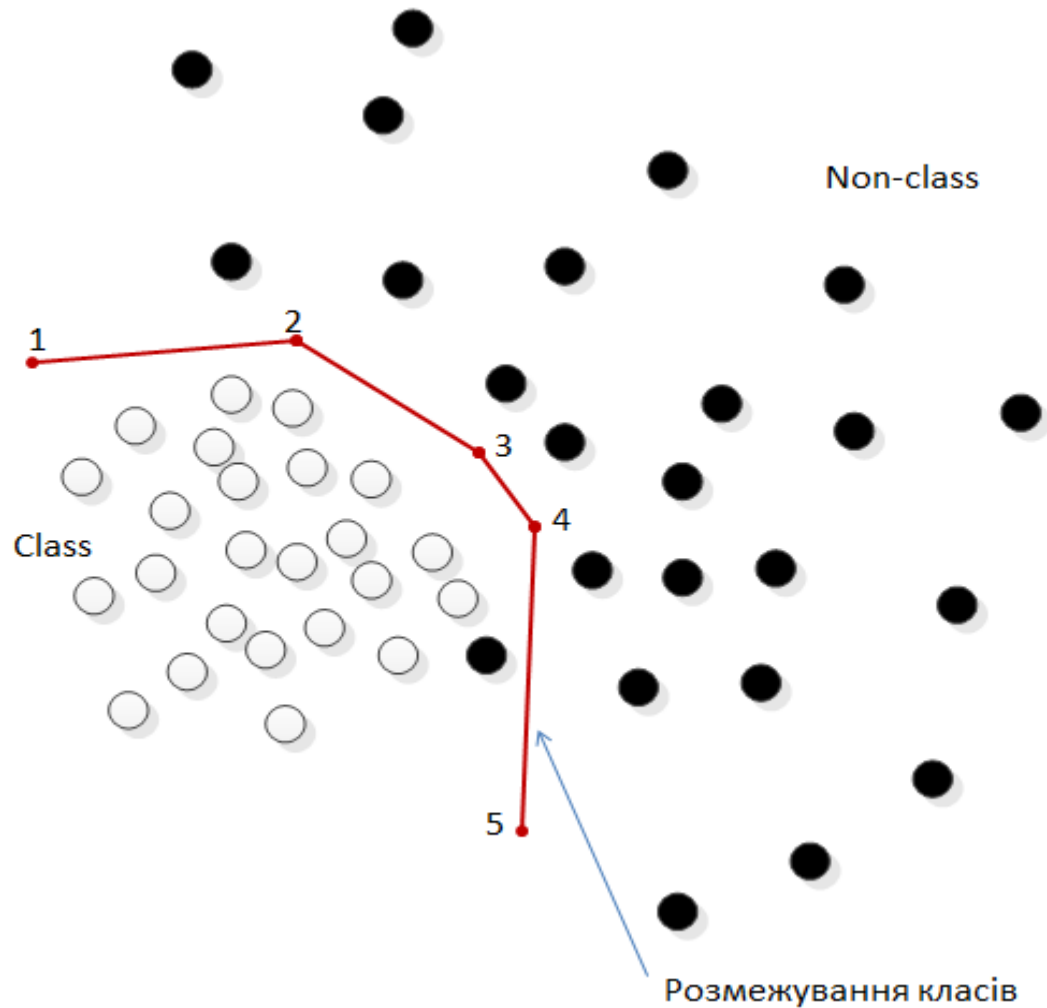


'ada-boost',
Точність 0,9679; Повнота 0,9063

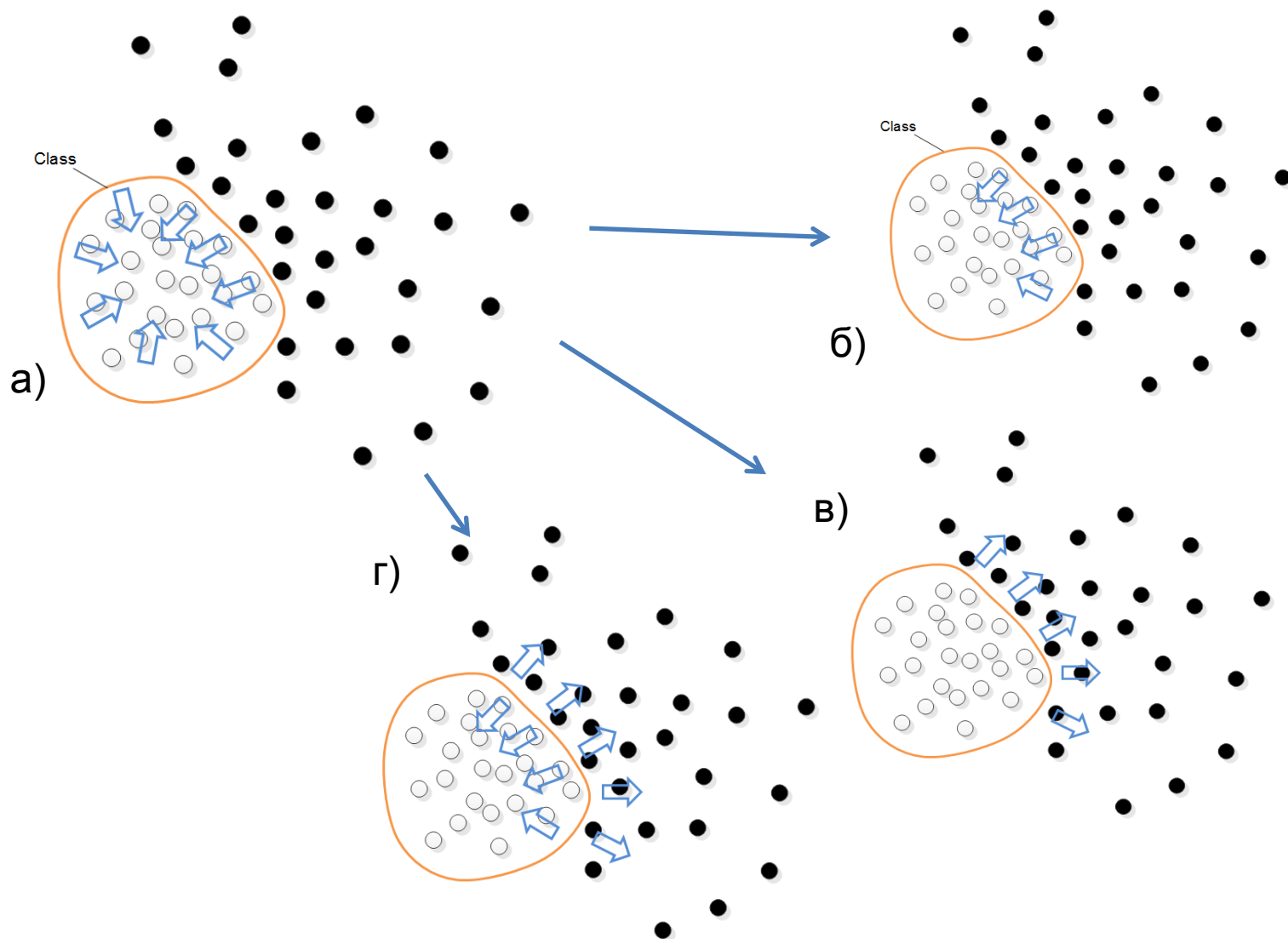


'ada-boost', 'random-forest'
Точність 0,9551; Повнота 0,9745,

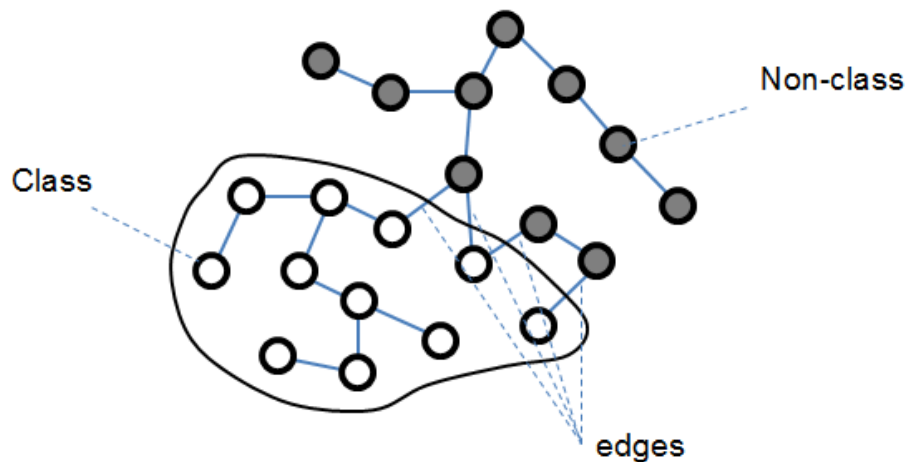
Кусково-лінійне розмежування класів побудоване людиною на редукованому просторі ознак



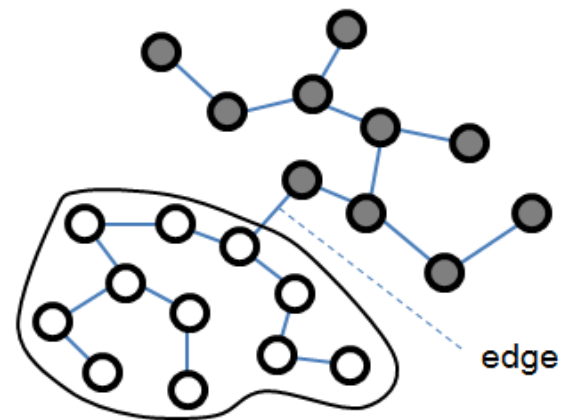
Види зміщення груп даних для визначення ознак, які мають найбільший вплив на розмежовування даних



Групування перехідних даних та ущільнення на основі приграничного розмежування



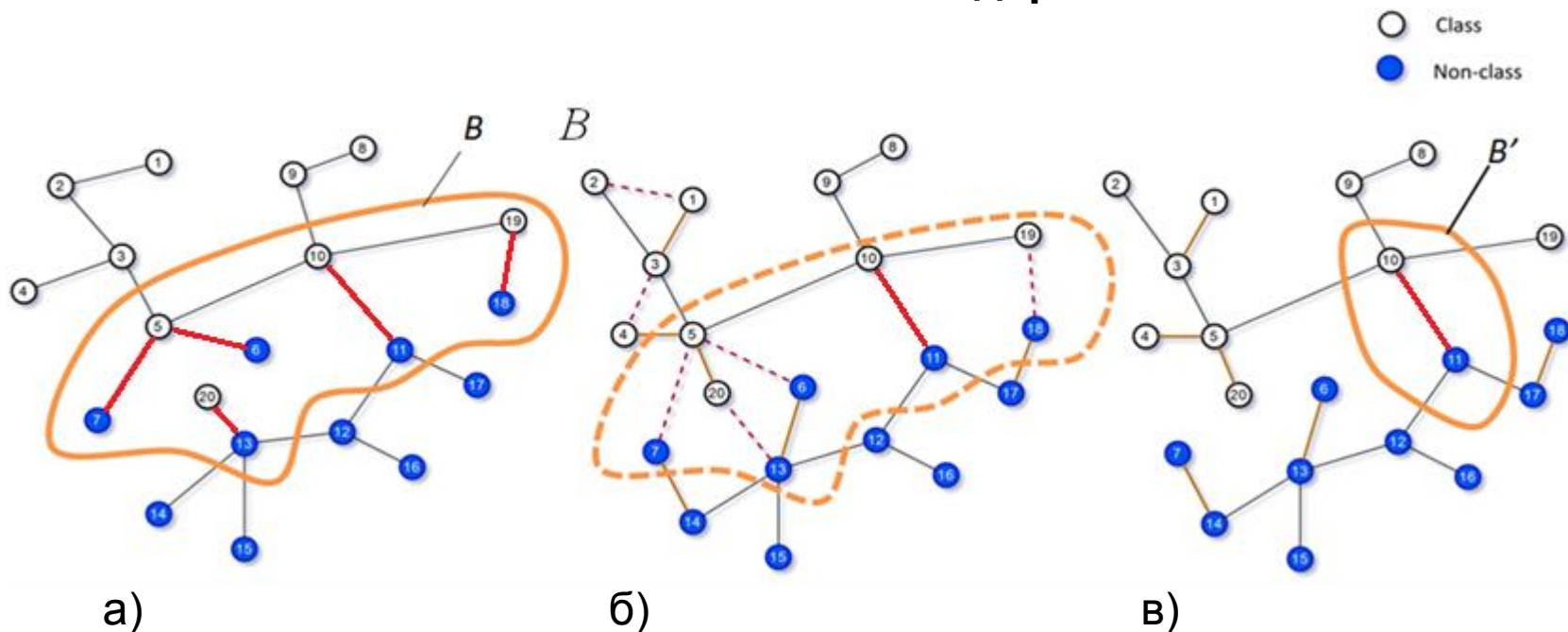
а)



б)

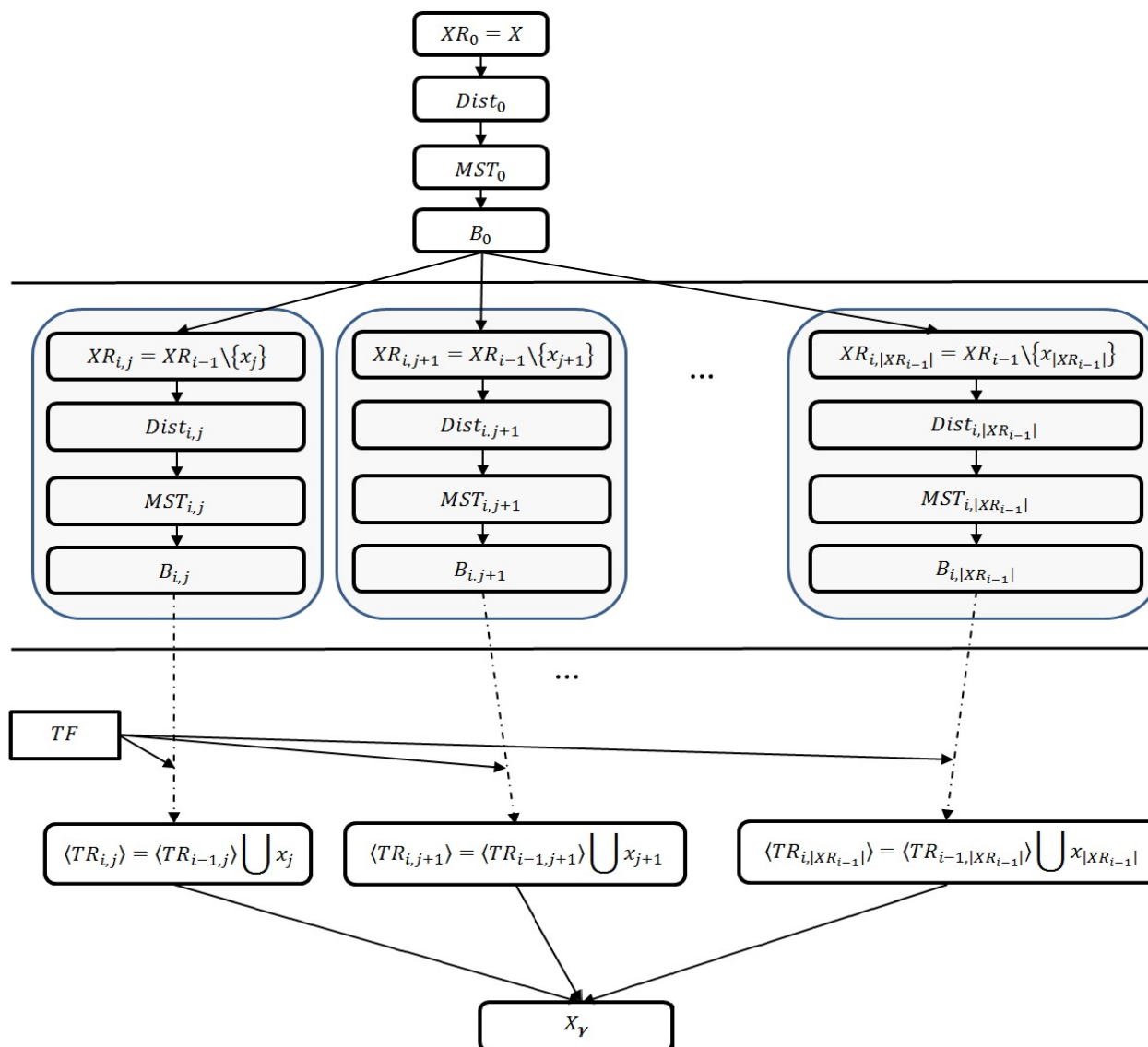
Зона перехідних даних D_{Trans} з MST і позначеними ребрами, які з'єднують дані класу D_{Class} із даними за межами класу $D_{NonClass}$: а) дані в початковому стані з позначеними ребрами з'єднання об'єктів D_{Class} і $D_{NonClass}$; б) дані перехідної зони після зміни ознак і збільшення компактності даних класу та формування ідеальних умов для просторового розташування даних класу, які відповідають наявності одного ребра з'єднання з даними за межами класу.

Ілюстрація етапів просторового зміщення з перетворенням мінімального остового дерева

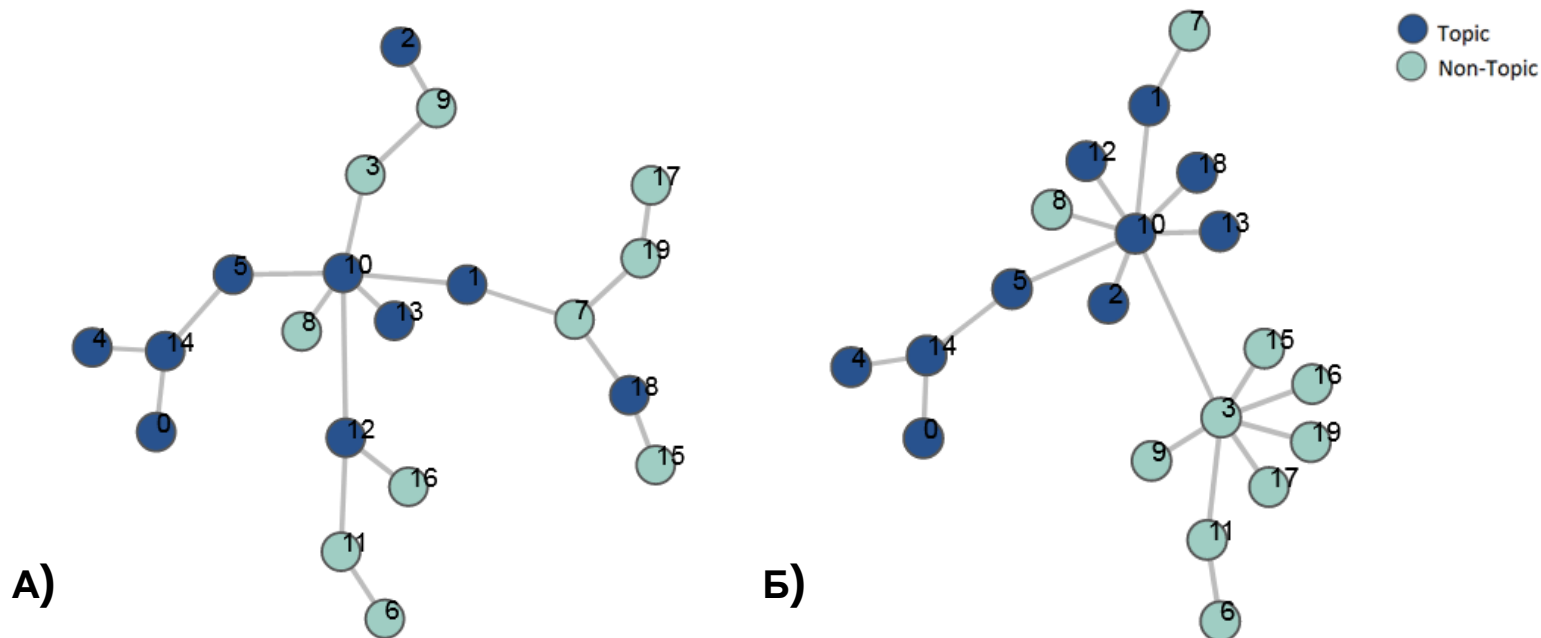


Етапи перетворення мінімального остовного дерева (MST) даних перехідної зони з виділенням двочасткового підграфу та його мінімізацією (у кращому випадку двочастковий граф має мінімально можливе значення кількості ребер рівне одиниці). B - початковий двочастковий підграф MST, B' - двочастковий підграф MST після просторового зміщення вершин. Пунктирна лінія позначає ребра MST, які зникають після перетворення дерева: а) початковий MST з двочастковим підграфом; б) перетворення MST; в) просторово трансформований MST з двочастковим підграфом.

Формування множин ознак у паралельній формі



Експериментальні дослідження визначення ознак найбільш впливових на роздільність перехідних даних (корпус Reuters-21578)



MST для початкового стану перехідних даних $n = 20$, побудованих з урахуванням лише спільних ознак 187. Кількість ребер дводольного графа 8, кількість всіх ознак 778

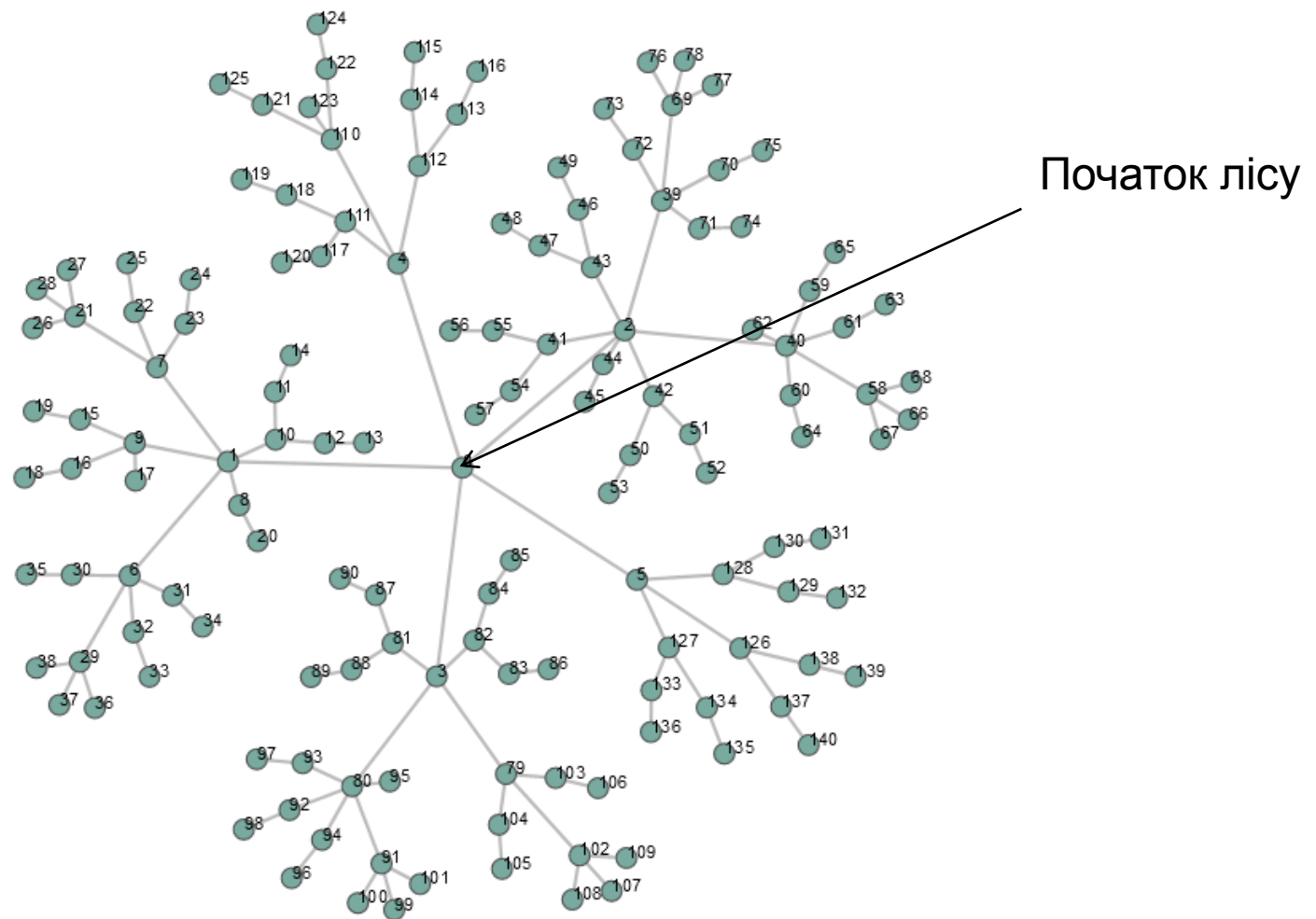
За цієї умови отримаємо три набори після зведення кортежей із спільними ознаками $\{372, 306, 522, 455\}$, $\{616, 306, 522, 455\}$, $\{455, 306, 522, 445\}$.

Підрахунок ознак на їх наявність у цих наборах.

522 $\rightarrow$ 3 306 $\rightarrow$ 3 455 $\rightarrow$ 3, 372 $\rightarrow$ 1, 616 $\rightarrow$ 1, 445 $\rightarrow$ 1

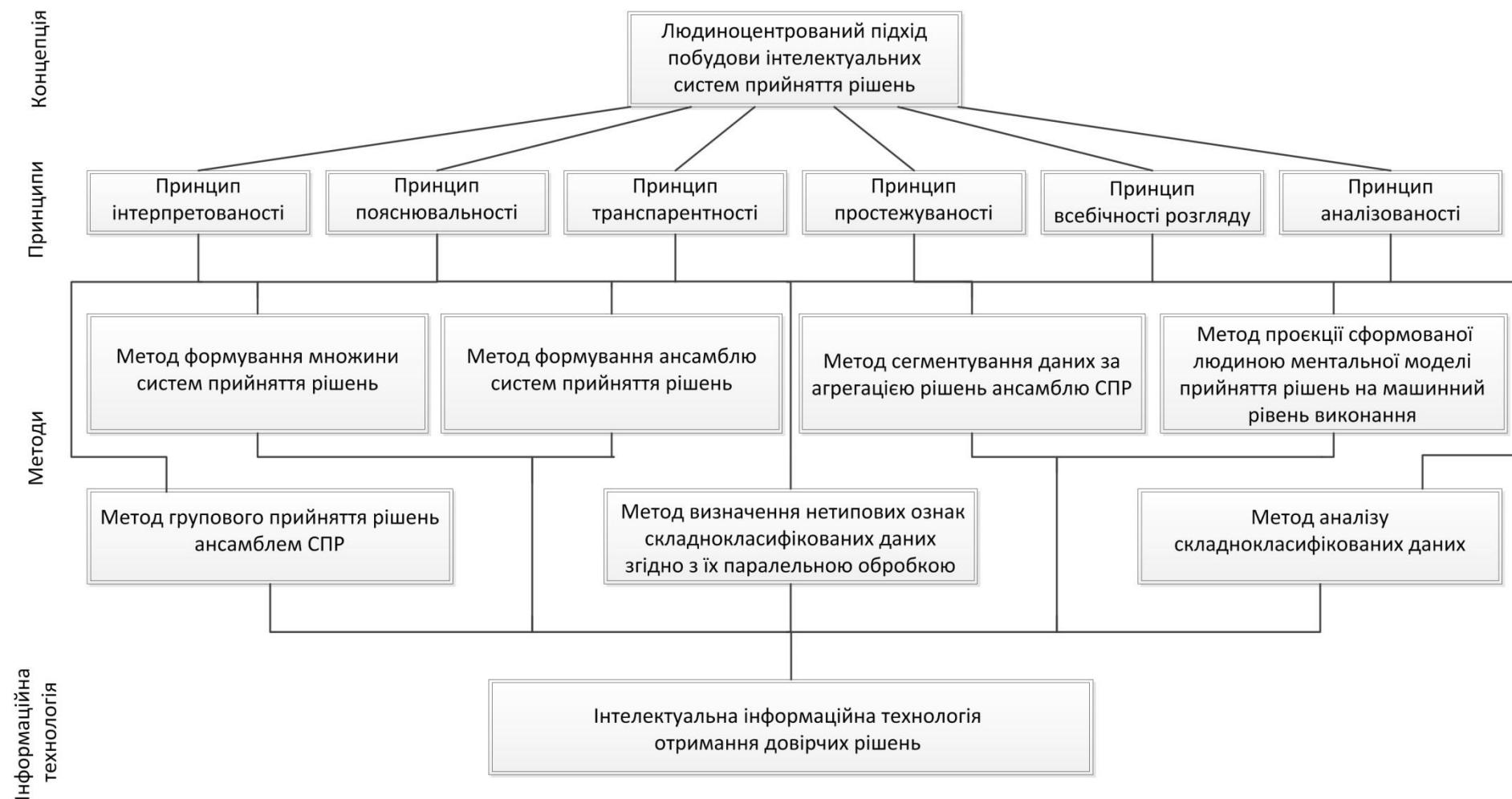
У кожній траєкторії наявні ознаки під номерами 522, 306, 455, що призводять до необхідної мінімізації. Це говорить про те, що ці ознаки можна вважати викидами. Після зміни ознак число ребер BG – 3.

Приклад лісу графів із траєкторій ознак як призводять до збільшення роздільності перехідних даних

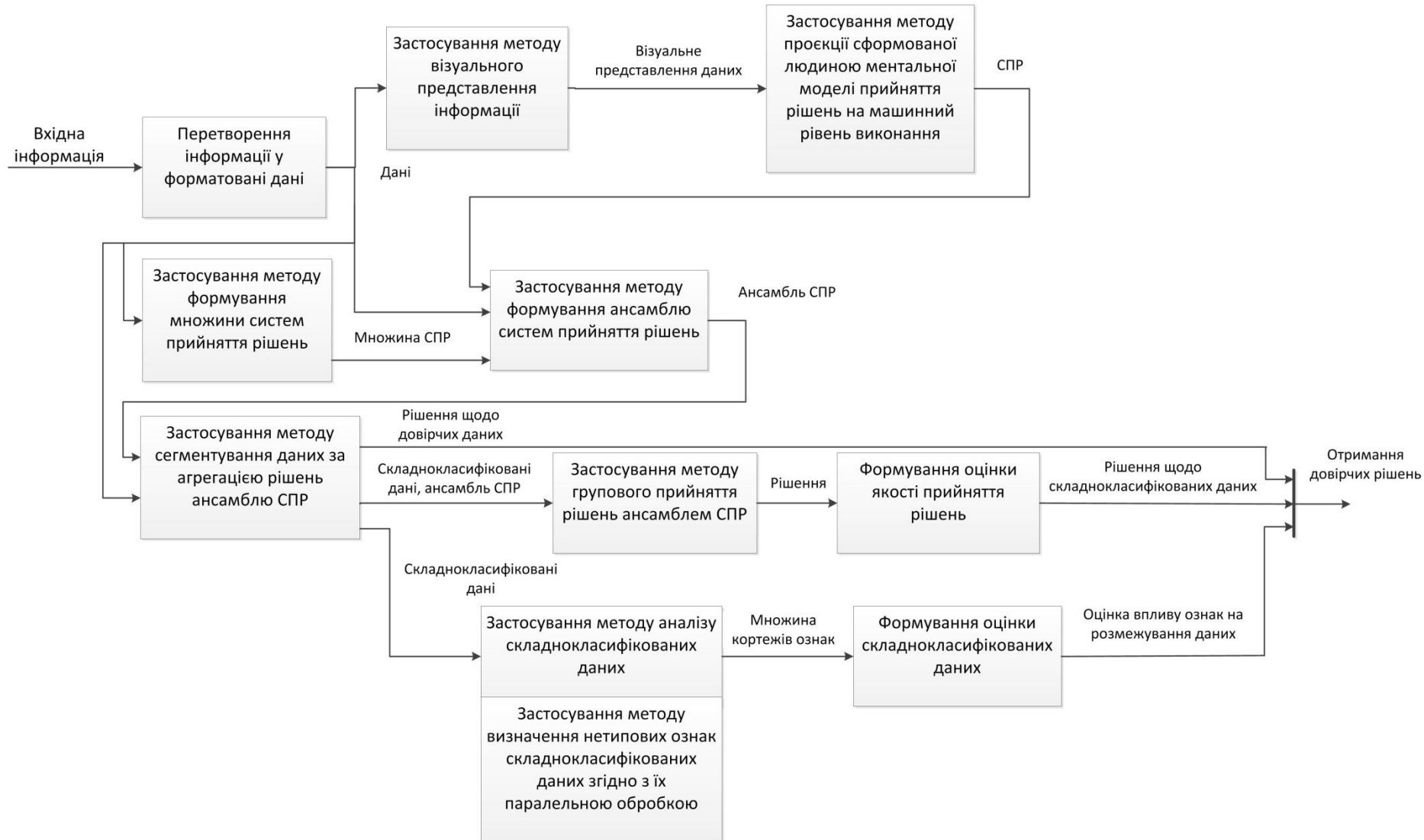


Графік траєкторій за сукупністю спільних ознак. Вершина з позначкою «0» вказує на початок графів лісу і не є ознакою

Структура формування інтелектуальної інформаційної технології отримання довірчих рішень



Структура взаємодії складових інтелектуальної інформаційної технології отримання довірчих рішень



Числові результати експериментальних досліджень

1. При високих показниках базових систем прийняття рішень покращення знаходиться для точності в межах 1–3.5 %, повноти 3–9 %. Відзначається загальне підвищення повноти над точністю, що пояснюється використанням методом групування та структурою ієрархії з орієнтуванням на врахування окремих рішень та узагальнення рішень груп.
2. Сегмент довірчих даних знаходиться в межах 76–86 % від загального обсягу даних.
3. Зменшення міжкласових ребер двочасткового графу 12.5 - 62.5 %, що відповідає покращенню розмежованості складнокласифікованих даних.
4. Частка найбільш впливових ознак на розмежування даних в сегменті складнокласифікованих даних становить близько 1.5 - 4.2 % від усіх ознак.