

ДЕРЖАВНИЙ  
СТАНДАРТ УКРАЇНИ

МІЖДЕРЖАВНИЙ  
СТАНДАРТ

---

Інформація та документація

**ОБСТЕЖЕННЯ ДОКУМЕНТА,  
ВСТАНОВЛЕННЯ ЙОГО ПРЕДМЕТА  
ТА ВІДБІР ТЕРМІНІВ ІНДЕКСУВАННЯ**

Загальна методика

ДСТУ 2395–2000 (ГОСТ 30671–99)  
(ISO 5963:1985)

БЗ № 3–2000/40

Информация и документация

**ОБСЛЕДОВАНИЕ ДОКУМЕНТА,  
УСТАНОВЛЕНИЕ ЕГО ПРЕДМЕТА  
И ОТБОР ТЕРМИНОВ ИНДЕКСИРОВАНИЯ**

Общая методика

ГОСТ 30671–99 (ИСО 5963–85)

*Видання офіційне*



ДСТУ 2395–2000  
(ГОСТ 30671–99)  
(ISO 5963:1985)

ДЕРЖАВНИЙ СТАНДАРТ УКРАЇНИ

---

Інформація та документація

**ОБСТЕЖЕННЯ ДОКУМЕНТА,  
ВСТАНОВЛЕННЯ ЙОГО ПРЕДМЕТА  
ТА ВІДБІР ТЕРМІНІВ ІНДЕКСУВАННЯ**

**Загальна методика**

*Видання офіційне*

## ПЕРЕДМОВА

- 1 РОЗРОБЛЕНО І ВНЕСЕНО ВАТ «Інститут прикладної інформатики»
- 2 ЗАТВЕРДЖЕНО І ВВЕДЕНО В ДІЮ наказом Держстандарту України від 27 листопада 2000 р. № 677
- 3 Цей стандарт відповідає міжнародному стандарту ISO 5963:1985 Documentation — Methods for examining documents, determining their subjects, and selecting indexing terms (Документація — Методи обстеження документів, установлення їхніх предметів і відбору термінів індексування)  
Ступінь відповідності — еквівалентний (eqv)
- 4 НА ЗАМІНУ ДСТУ 2395–94
- 5 РОЗРОБНИКИ: А. О. Стогній, чл.-кор. НАН України; В. М. Дріяньський, канд.фіз.-мат.наук; С. М. Альошкіна; Т. Н. Давиденко; О. Ю. Дегтярьова; В. В. Діденко; О. М. Коренга

## ЗМІСТ

	С.
1 Сфера застосування .....	1
2 Визначення .....	2
3 Призначення і процедура індексування .....	2
4 Обстеження документа .....	3
5 Ідентифікація понять .....	3
6 Відбір термінів індексування .....	5
7 Контроль якості індексування .....	5
Додаток А Блок-схема процедури індексування з використанням тезауруса .....	7

ДСТУ 2395–2000  
(ГОСТ 30671–99)  
(ISO 5963:1985)

## ДЕРЖАВНИЙ СТАНДАРТ УКРАЇНИ

---

ІНФОРМАЦІЯ ТА ДОКУМЕНТАЦІЯ

### ОБСТЕЖЕННЯ ДОКУМЕНТА, ВСТАНОВЛЕННЯ ЙОГО ПРЕДМЕТА ТА ВІДБІР ТЕРМІНІВ ІНДЕКСУВАННЯ

Загальна методика

ИНФОРМАЦИЯ И ДОКУМЕНТАЦИЯ

### ОБСЛЕДОВАНИЕ ДОКУМЕНТА, УСТАНОВЛЕНИЕ ЕГО ПРЕДМЕТА И ОТБОР ТЕРМИНОВ ИНДЕКСИРОВАНИЯ

Общая методика

INFORMATION AND DOCUMENTATION

### EXAMINING THE DOCUMENT, DETERMINING IT'S SUBJECT, AND SELECTING INDEXING TERMS

General methods

---

Чинний від 2001–07–01

## 1 СФЕРА ЗАСТОСУВАННЯ

Цей стандарт поширюється на процес індексування документа і встановлює загальну методику виконання початкових його стадій, як-от:

- обстеження документа та встановлення його предмета;
- ідентифікація понять, що становлять предмет;
- відбір відповідних термінів індексування.

Цей стандарт не пов'язаний з будь-якою конкретною системою індексування: передкоординатною (порядок термінів чи класифікаційних індексів устанавлюється наперед) чи посткоординатною (порядок термінів чи класифікаційних індексів відсутній).

Цей стандарт передусім орієнтований на системи індексування дескрипторного типу (поняття виражають термінами інформаційної мови, що контролюються, наприклад, тезаурусом), але він також придатний до систем індексування класифікаційного типу (поняття виражають індексами класифікаційної системи).

Цей стандарт слід застосовувати в усіх ситуаціях індексування за участю особи, що здійснює індексування (далі — індексатор), під час аналізування документів і пошукових запитів користувачів, складання рефератів, підготування предметних покажчиків для забезпечення уніфікації технології індексування документів, передусім, в організаціях, що обмінюються бібліографічною інформацією.

Цей стандарт не придатний у випадках автоматичного індексування, коли терміни, що зустрічаються в тексті, поєднуються в множини чи класи на підставі формально заданих критеріїв, наприклад, частоти появи і/або близькості в тексті.

## 2 ВИЗНАЧЕННЯ

У цьому стандарті використано такі терміни й визначення:

**2.1 Документ** — записана інформація, яку можна розглядати як одиницю в ході здійснення інформаційної діяльності.

*Примітка.* Це визначення стосується не тільки рукописних матеріалів чи друківаних на папері чи в мікроформах (книжки, журнали, діаграми, карти), але й недруківаних (машинозчитних записів, фільмів, звукових записів), а також тривимірних об'єктів чи реалій.

**2.2 Записана інформація** — інформація, що зберігається за допомогою носія даних.

**2.3 Інформаційна діяльність** — постійне та систематичне збирання та оброблення записаної інформації з метою її зберігання, пошуку використання чи пересилання, що виконуються особою чи організацією.

**2.4 Поняття** — будь-яка одиниця думки.

*Примітка.* Семантичний зміст поняття можна передавати комбінацією інших понять, склад яких залежить від мови чи культури.

**2.5 Предмет** — поняття чи комбінація понять, що відбивають тему документа.

**2.6 Індекссування** — процес подання результатів аналізу документа елементами інформаційної чи природної мови, як правило, з метою полегшення його пошуку.

**2.7 Термін індексування** — представлення поняття як елемента інформаційної мови чи терміна природної мови (переважно іменника чи іменного словосполучення), або класифікаційного індексу.

**2.8 Інформаційна мова** — формалізована мова, призначена для характеристики даних чи змісту документів із метою забезпечення їхнього зберігання та пошуку.

**2.9 Преференційний термін** — термін, вибраний як дескриптор із множини еквівалентних термінів.

**2.10 Дескриптор** — термін тезауруса, який може бути використаний для вираження поняття в документі чи запиті з метою інформаційного пошуку.

**2.11 Тезаурус** — контрольований словник термінів із зафіксованими семантичними відношеннями, який охоплює одну чи більше спеціальних галузей знань.

**2.12 Класифікаційна система** — інформаційна мова, призначена для структурного подання документів чи даних за допомогою класифікаційних індексів і відповідних термінів із метою забезпечення реалізації класифікаційного предметного підходу з використанням, у разі необхідності, абеткового покажчика.

**2.13 Класифікаційний індекс** — подання класу, отримане внаслідок застосування нотації деякої класифікаційної системи.

**2.14 Нотація** — система позначок (кодів) і правил їхнього застосування, що використовується для подання класів та відношень між ними.

*Примітка.* З метою структурування можуть бути використані спеціальні символи, наприклад, знаки пунктуації.

**2.15 Покажчик** — абетковий чи систематичний перелік предметів із посиланням на позицію кожного предмета в документі чи сукупності документів.

**2.16 Релевантність** — характеристика ступеня відповідності змісту документа, знайденого в результаті інформаційного пошуку, змісту інформаційного запита.

## 3 ПРИЗНАЧЕННЯ І ПРОЦЕДУРА ІНДЕКСУВАННЯ

**3.1 Індекссування** полягає в ідентифікації змісту документа з метою його наступного відшукування. Індекссування не передбачає описування документа як фізичної одиниці, хоча деякі параметри щодо форми документа, його вихідних даних тощо можуть бути занесені до предметного покажчика, якщо інформація такого роду дозволить користувачеві точніше визначити, чи є цей документ релевантним його запиту.

3.2 Під час індексування відбувається видобування понять із документів шляхом інтелектуального аналізу і перетворення їх у терміни індексування. Як аналіз, так і перетворення слід здійснювати за допомогою таких засобів індексування, як тезауруси та класифікаційні системи.

3.3. Процедура індексування містить такі три взаємопов'язані процеси:

- обстеження змісту документа та встановлення його предмета;
- ідентифікація основних понять, що становлять предмет документа;
- відбір відповідних термінів індексування для вираження виявлених понять засобами вибраної інформаційної мови.

Рекомендації та вимоги щодо виконання цих процесів, а також щодо якості індексування розглянуто в розділах 4—7.

## 4 ОБСТЕЖЕННЯ ДОКУМЕНТА

4.1 Результатом обстеження документа має бути встановлення його предмета.

Докладність обстеження документа значною мірою залежить від його фізичної форми. Слід розрізняти такі два випадки:

- індексування друкованих документів, що становлять звичайне явище для бібліотек та інформаційних центрів, де фонд складається здебільшого з монографій, часописів, звітів, матеріалів конференцій тощо;
- індексування аудіовізуальних творів чи фонограм.

4.2 Найповніше розуміння друкованих документів досягається в результаті докладного прочитання всього тексту. Таке повне прочитання часто неможливо реалізувати, і воно не завжди є обов'язковим, однак індексатор повинен гарантувати, що жодної корисної інформації не втрачено.

Обстежуючи документ, індексатор повинен звернути особливу увагу на такі важливі частини тексту:

- назва;
- реферат (якщо є);
- зміст;
- вступ, початкові фрази глав та параграфів, висновки;
- ілюстрації, діаграми, таблиці та супровідні надписи;
- слова чи фрази, підкреслені чи виділені яким-небудь іншим чином (наприклад надруковані іншим шрифтом).

Індексування тільки на основі назви не рекомендоване. Реферат (якщо він є) також не слід розглядати як заміну тексту документа. Як назви, так і реферати можуть бути невідповідними тексту і не можуть вважатися надійним джерелом, необхідним індексатору.

4.3 У випадку аудіовізуальних творів чи фонограм доводиться враховувати той факт, що, наприклад, аудіовізуальні, візуальні та звукові носії потребують додаткових технічних процедур. На практиці не завжди є можливість обстеження самого документа (наприклад прокручування плівки). Індексування в такому випадку звичайно виконується на основі назви і/або стислого опису, хоч індексатор повинен мати можливість переглянути чи прослухати інформацію безпосередньо з носія, якщо письмовий опис виявиться невідповідним або неточним.

## 5 ІДЕНТИФІКАЦІЯ ПОНЯТЬ

5.1 Після обстеження документа індексатор повинен перейти до систематизованої ідентифікації тих понять, які є суттєвими елементами предмета документа.

Зацікавлені організації повинні встановити контрольні списки тих параметрів, які визнано важливими у відповідній предметній галузі. У наведених нижче запитаннях ілюструються загальні параметри, які мають бути включені до такого контрольного списку.

Чи описується в документі об'єкт діяльності?

Чи містить предмет активне поняття (наприклад дію, операцію, процес тощо)?

Чи ідентифікований об'єкт діяльності?

Чи описується в документі суб'єкт дії?

Чи має документ посилання на конкретні засоби виконання дії (наприклад спеціальні інструменти, технології чи методи)?

Чи розглядалися ці параметри в контексті конкретного місцеположення чи середовища?

Чи ідентифіковані які-небудь залежні або незалежні змінні величини?

Чи розглядався предмет у спеціальному аспекті, як правило, не пов'язаному з галуззю дослідження (наприклад соціологічне дослідження релігії)?

Наведені запитання є прикладами загальних параметрів, які можуть бути застосовані в будь-якій предметній галузі. У кожній спеціальній дисципліні може виявитися потрібне формулювання інших запитань.

5.2 Індиксатор не повинен обов'язково ідентифікувати всі поняття, що становлять предмет документа. Відбір понять залежить від мети індексування.

Ідентифікація понять може також залежати від виду документа. Наприклад, результати індексування на основі текстів книжок, статей часописів тощо найчастіше відрізняються від результатів індексування на основі рефератів чи коротких оглядів.

Докладність ідентифікації понять пов'язана з такими характеристиками індексування як чіткість та специфічність.

5.3 Вичерпність індексування залежить від числа параметрів (5.1), поданих поняттями, явленими індиксатором.

5.3.1 Індиксатор повинен бути спроможний ідентифікувати всі поняття документа, що мають потенційну значимість для користувачів тієї чи іншої інформаційної системи. Іноді дві чи більше тем у межах відповідної предметної галузі розглядаються незалежно в одному й тому самому документі. Вони повинні бути опрацьовані окремо і, якщо необхідно, різними спеціалістами.

5.3.2 Межі відповідної предметної галузі не можна розглядати надто звужено. Потрібно виховувати той факт, що з розвитком інформаційних мереж результати індексування, проведеного спочатку для однієї групи користувачів (наприклад учених чи технологів), можуть застосовуватися іншими групами користувачів (наприклад економістами). З огляду на це індиксаторам, наприклад, наукової та технічної літератури рекомендовано не ігнорувати інші аспекти предмета, наприклад соціальний та економічний.

5.3.3 У процесі відбору понять головним критерієм завжди повинна бути потенційна значимість поняття як елемента вираження предмета документа і як параметра для пошуку документа. Відбираючи поняття, індиксатор повинен враховувати, наскільки це можливо, запитання, які надходять до тієї чи іншої інформаційної системи. Фактично цей критерій відбиває нову функцію індексування. З огляду на це індиксатор повинен:

— відібрати поняття, що є найпридатнішими, з погляду деякої заданої групи користувачів пам'ятаючи при цьому про завдання індексування;

— змінити, якщо необхідно, процедуру індексування, виходячи з результатів зворотного зв'язку після опрацювання запитів. Проте така зміна не повинна суперечити структурі й логіці мови індексування.

5.3.4 Не встановлюється жодних обмежень щодо кількості термінів індексування, отриманих за певний період, приписаних одному документу. Кількість понять визначається тільки кількістю інформації, що міститься в документі, з урахуванням завдань індексування. Встановлення будь-яких обмежень щодо кількості понять може призвести до деякої втрати об'єктивності в процесі індексування і до втрати інформації, яка могла б бути корисною під час пошуку. Якщо в рамках якоїсь організації виникає потреба в обмеженні кількості термінів індексування, відбір понять має визначатися міркуваннями індиксатора щодо ролі того чи іншого поняття у вираженні предмету документа.

5.4 Специфічність індексування визначається тим ступенем точності, з яким певне поняття документа подається засобами мови індексування. Специфічність втрачається, коли поняття подається терміном більш загального значення.

Поняття повинні бути ідентифіковані з максимально можливим ступенем специфічності. За певних обставин більш загальні поняття можуть бути прийнятнішими, наприклад у таких випадках:



— коли високий ступінь специфічності, на думку індексатора, може негативно вплинути на характеристики системи індексування. Наприклад, індексатор може вирішити замінити специфічні найменування моделей устаткування більш загальними термінами, наприклад найменуванням виробника чи найменуванням сімейства моделей, особливо, якщо ці поняття знаходяться на межі відповідної предметної галузі;

— якщо індексатор вважає, що деяка ідея розкрита не досить повно або зазначена автором лише побіжно, і приписує відповідному поняттю низьку вагу.

## 6 ВІДБІР ТЕРМІНІВ ІНДЕКСУВАННЯ

6.1 Під час подання виявлених понять термінами індексування індексатор повинен дотримуватися таких правил (див. також додаток А):

— поняття, подані в інформаційній мові, повинні бути виражені відповідними преференційними термінами (тобто дескрипторами);

— терміни, що виражають нові поняття, повинні перевірятися щодо точності та прийнятності із залученням таких засобів: словників та енциклопедій, визнаних авторитетними в цій галузі; тезаурусів; класифікаційних систем.

Бажані також консультації зі спеціалістами відповідної предметної галузі, особливо з тими, хто має досвід у сфері індексування.

6.2 Індексатор повинен бути обізнаним із правилами застосування цих засобів індексування. Зокрема, він повинен усвідомлювати, що зазначені засоби можуть накладати певні обмеження. Наприклад, заданий наперед перелік предметних заголовків або склад розділів класифікаційної системи можуть не дозволити точно виразити поняття, зустрінуті в документі. Якщо поняття виражені класифікаційними індексами, слід розуміти, що ці індекси звичайно визначають ширший чи вузьчий контекст (тобто основний клас), який може й не зовсім відповідати обстежуваному документу.

6.3 Якщо інформаційна мова включає тезаурус, кількість термінів, приписаних документу, може бути зменшена без втрат, тому що родові та інші відношення, що існують априорі, можуть бути встановлені безпосередньо на підставі тезауруса. У випадку використання тезауруса для вираження певного поняття слід добирати найбільш спеціальний термін.

6.4 У деяких системах індексування використовуються індикатори ролі, зв'язку, ваги тощо. Індексатор повинен бути обізнаний з усіма спеціальними правилами, пов'язаними з цими механізмами.

6.5 На практиці індексатор часто зустрічає поняття, які відсутні в наявному тезаурусі чи класифікаційній системі. Залежно від використовуваної системи ці поняття можуть бути опрацьовані по-різному, наприклад:

— вони можуть бути виражені термінами чи дескрипторами, які відразу вводяться в інформаційну мову;

— вони можуть бути тимчасово виражені більш загальними термінами, а нові поняття можуть бути запропоновані як потенційні для пізнішого поповнення інформаційної мови.

## 7 КОНТРОЛЬ ЯКОСТІ ІНДЕКСУВАННЯ

7.1 Якість індексування залежить від таких чинників:

— кваліфікація й досвід індексатора;

— якість засобів індексування.

Якість індексування значною мірою визначається несуперечливістю індексування, яка полягає в тому, що склад термінів індексування, приписаних деякому документу, і досягнуті під час індексування рівні вичерпності та специфічності є в ідеалі незалежними від індексатора і залишаються незмінними протягом функціонування конкретної системи індексування. На практиці не завжди можливо досягти ідеалу несуперечливості, але прагнення до нього є важливим чинником, що впливає на характеристики системи індексування, особливо в разі обміну інформацією між різними організаціями в мережі.

7.2 Обов'язковою умовою для досягнення несуперечливості індексування є повна неупередженість з боку індексатора в процесі ідентифікації понять і вибору термінів індексування. У разі коли є велика група індексаторів, що працюють у різних місцях, наприклад у децентралізованих системах, для досягнення несуперечливості індексування рекомендовано введення етапу централізованої перевірки зі зворотним зв'язком з індексаторами.

7.3 Для забезпечення достатньої якості індексування індексатор повинен мати відповідні знання в тій предметній галузі, до якої відносяться обстежувані документи. Він повинен розуміти терміни, що зустрічаються в документах, а також правила застосування конкретної інформаційної мови.

У разі індексування чужомовних документів індексатор повинен володіти відповідною мовою або мати можливість звернутися по допомогу до відповідних фахівців.

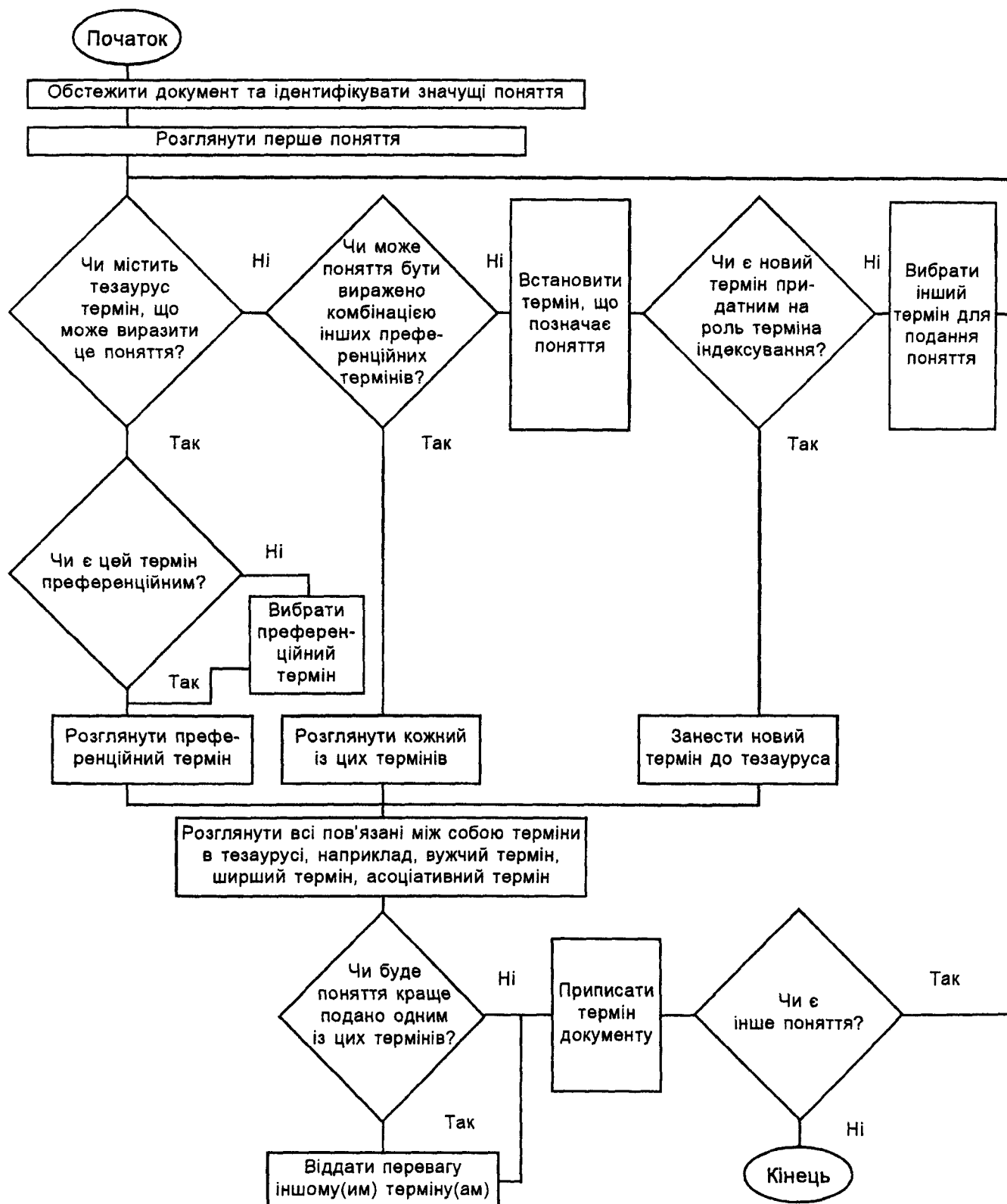
7.4 Підвищенню якості індексування сприятимуть контакти індексаторів із користувачами. Наприклад, з метою виявлення хибних термінологічних комбінацій, що можуть породжувати нерелевантну видачу під час інформаційного пошуку.

7.5 Якість індексування залежить також від гнучкості застосовуваної інформаційної мови. Бажано, щоб вона дозволяла вводити нові терміни та вносити зміни в термінологію.

7.6 Якість індексування слід перевіряти, де це можливо, шляхом аналізу результатів інформаційного пошуку, наприклад, із використанням коефіцієнтів повноти й точності.

ДОДАТОК А  
(довідковий)

**БЛОК-СХЕМА ПРОЦЕДУРИ ІНДЕКСУВАННЯ  
З ВИКОРИСТАННЯМ ТЕЗАУРУСА**



УДК 002:001.815:006.354

01.140.20

ТІ

**Ключові слова:** загальна методика, інформація, документація, індексування, предмет документа, термін індексування, інформаційна мова, тезаурус, класифікаційна система, інформаційна діяльність.

---