

**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)
(180 5963:1985)**

Информация и документация

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

Общая методика

ГОСТ 30671-99 (ИСО 5963-85)

Видання офіційне

ДЕРЖСТАНДАРТ УКРАЇНИ
Київ

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

ДЕРЖСТАНДАРТ УКРАЇНИ
Київ

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 ІЗО 5963:1985 Оситепіаііоп — Меіпосіз тог ехатіпід сіоситепіз, а!ет.егтіпід іпеіг зібіесіз, апсі зеїестіпд іпсіехіпд іегтз (Документація — Методи обстеження документів, установлення їхніх предметів і відбору термінів індексування) Ступінь відповідності — еквівалентний (ецч)
- 4 НА ЗАМІНУ ДСТУ 2395-94
- 5 РОЗРОБНИКИ: **А. О. Стогній**, чл.-кор. НАН України; **В. М. Дряньський**, 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

EXAMINATION OF THE DOCUMENT, DETERMINATION OF ITS SUBJECT, AND SELECTION OF TERMS FOR INDEXING

Общая методика

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 встановлюється 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.

Видання офіційне

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 Індексвання — 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 Індексвання полягає в ідентифікації змісту документа з метою його наступного відшукування. Індексвання 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 про читання всього тексту. Таке повне прочитання часто неможливо реалізувати, і воно не завжди є обов'язковим, однак індексатор повинен гарантувати, що жодної корисної інформації не впущено.

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

- назва;
- реферат (якщо є);
- зміст;
- вступ, початкові фрази глав та параграфів, висновок;
- 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 потрібним формулювання інших запитань.

5.2 Індексатор 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 вичерпність та специфічність.

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

5.3.1 Індексатор 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 враховувати 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 індексаторам, 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, індексатор повинен враховувати, наскільки це можливо, запитання, які надходять до тієї чи іншої інформаційної системи. Фактично цей критерій відбиває основну функцію індексування. З огляду на це індексатор 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, отже й понять, 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 якої-небудь організації виникає потреба в обмеженні кількості термінів індексування, відбір понять має визначатися міркуваннями індексатора щодо ролі того чи іншого поняття у вираженні предмета 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.

ДСТУ 2395-2000

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.

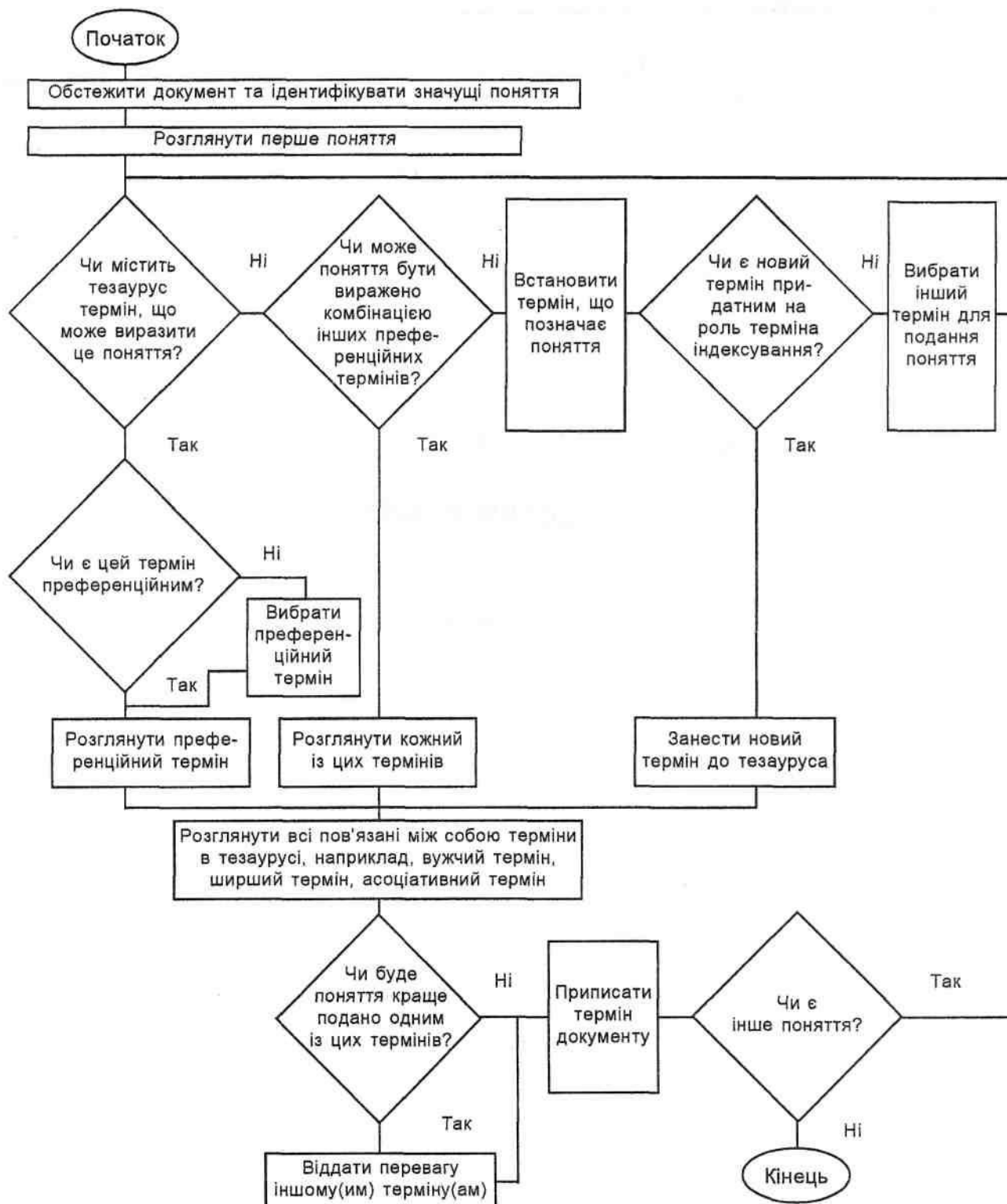
7.4 Підвищенню якості індексування сприятимуть контакти індексаторів із користувачами. Наприклад, з метою виявлення хибних термінологічних комбінацій, що можуть породжувати нерелевантну видачу під час інформаційного пошуку.

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



ДСТУ 2395-2000

УДК 002:001.815:006.354

01.140.20

T62

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

ГОСТ 30671-99
(ИСО 5963-85)

МЕЖГОСУДАРСТВЕННЫЙ СТАНДАРТ

Информация и документация

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

Общая методика

Издание официальное

Межгосударственный совет по стандартизации,
метрологии и сертификации

МЕЖГОСУДАРСТВЕННЫЙ СТАНДАРТ

ИНФОРМАЦИЯ И ДОКУМЕНТАЦИЯ

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

Общая методика

ТРОРМАТЮЫ АМР йОСИМЕНТАТЮМ

ЕХАМИЧШСВ ТНЕ ООС1ШЕ1ЧТ,
ОЕТЕРМ1МЫЮ 1гз ЗЫБ^СТ,
А^ ЗЕЬЕСТ^Б 1ШЕХ1ЫС ТЕКМ5

Оепера! tel!гюс1з

Дата введения 2001-07-01

1 ОБЛАСТЬ ПРИМЕНЕНИЯ

Настоящий стандарт распространяется на процесс индексирования документа и устанавливает общую методику выполнения начальных его стадий:

- обследование документа и установление его предмета;
- идентификация понятий, составляющих предмет;
- отбор соответствующих терминов индексирования.

Настоящий стандарт не связан с какой-либо конкретной системой индексирования: предкоординатной (порядок терминов или кодов классов устанавливается заранее) или посткоординатной (порядок терминов или кодов классов отсутствует).

Настоящий стандарт, в первую очередь, ориентирован на системы индексирования дескрипторного типа (понятия выражают терминами информационного языка, контролируемые, например, тезаурусом), но он также применим в системах индексирования классификационного типа (понятия выражают кодами классов классификационной системы).

Настоящий стандарт следует применять во всех ситуациях индексирования с участием лица, осуществляющего индексирование (далее — индексатор), при анализе документов, пользовательских поисковых запросов; при составлении рефератов; при подготовке предметных указателей для обеспечения унификации технологии индексирования документов, в первую очередь, в организациях, обменивающихся библиографической информацией.

Настоящий стандарт не пригоден в случаях автоматического индексирования, когда термины, встречающиеся в тексте, объединены во множества или классы на основании формально заданных критериев, например частоты встречаемости и/или близости в тексте.

2 ОПРЕДЕЛЕНИЯ

В настоящем стандарте использованы следующие термины и определения. 2.1 Документ — записанная информация, которую можно рассматривать как единицу в ходе осуществления информационной деятельности.

Издание официальное

Примечание — Это определение относится не только к рукописным или печатным материалам на бумаге или в виде микроформ (книгам, журналам, диаграммам, картам), но и к материалам непечатного характера (машиночитаемым записям, фильмам, звуковым записям), а также к трехмерным объектам или реалиям.

2.2 Записанная информация — информация, хранимая с помощью носителя данных.

2.3 Информационная деятельность — постоянный и систематический сбор и обработка записанной информации с целью ее сохранения, поиска, использования или пересылки, выполняемые каким-либо лицом или организацией.

2.4 Понятие — любая единица мысли.

Примечание — Семантическое содержание понятия можно передавать комбинацией других понятий, состав которых зависит от языка или культуры.

2.5 Предмет — понятие или комбинация понятий, отражающих тему документа.

2.6 Индексирование — процесс представления результатов обследования содержания документа элементами информационного языка или естественного языка, как правило, с целью облегчения его поиска.

2.7 Термин индексирования — представление понятия в форме элемента информационного языка или термина естественного языка (преимущественно имени существительного или именного словосочетания), или кода класса.

2.8 Информационный язык — формализованный язык, предназначенный для характеристики данных или содержания документов с целью обеспечения их хранения и поиска.

2.9 Предпочтительный термин — термин, выбранный в качестве дескриптора из множества эквивалентных терминов.

2.10 Дескриптор — термин тезауруса, который может быть использован для представления понятия в документе или запросе в целях информационного поиска.

2.11 Тезаурус — контролируемый словарь терминов с зафиксированными семантическими отношениями, который охватывает одну или более специальных областей знания.

2.12 Классификационная система — информационный язык, предназначенный для структурного представления документов или данных с помощью кодов классов и соответствующих терминов в целях обеспечения реализации классификационного предметного подхода с использованием, в случае необходимости, алфавитного указателя.

2.13 Код класса — представление класса, полученное в результате применения нотации некоторой классификационной системы.

2.14. Нотация — система обозначений (кодов) и правил их применения, используемая для представления классов и отношений между ними.

Примечание — В целях структуризации можно использовать специальные символы, например знаки пунктуации.

2.15 Указатель — алфавитный или систематический перечень предметов со ссылками на позицию каждого предмета в документе или совокупности документов.

2.16 Релевантность — характеристика степени соответствия содержания документа, найденного в результате информационного поиска, содержанию информационного запроса.

3 НАЗНАЧЕНИЕ И ПРОЦЕДУРА ИНДЕКСИРОВАНИЯ

3.1 Индексирование состоит в идентификации смысла документа с целью его последующего отыскания. Индексирование не предусматривает описания документа как физической единицы, хотя некоторые параметры, касающиеся формы документа, его выходных данных и т. п. могут быть внесены в предметный указатель, если эта информация позволит пользователю более точно определить, является ли документ релевантным его запросу.

3.2 В ходе индексирования происходит извлечение понятий из документов путем интеллектуального анализа и преобразование их в термины индексирования. Как анализ, так и преобразование следует осуществлять с помощью таких средств индексирования, как тезаурусы и классификационные системы.

3.3 Процедура индексирования состоит из следующих трех взаимосвязанных процессов: — обследование содержания документа и установление его предмета;

- идентификация основных понятий, составляющих предмет документа;
- отбор соответствующих терминов индексирования для представления выявленных понятий средствами выбранного информационного языка.

Рекомендации и требования к выполнению этих процессов, а также к качеству индексирования рассматриваются в разделах 4—7.

4 ОБСЛЕДОВАНИЕ ДОКУМЕНТА

4.1 Результатом обследования документа должно быть установление его предмета.

Тщательность обследования документа в значительной мере зависит от его физической формы. Следует различать следующие два случая:

- индексирование печатных документов, представляющих обычный случай в библиотеках и информационных центрах, где фонд состоит в основном из монографий, журналов, отчетов, материалов конференций и т.п.;
- индексирование аудиовизуальной продукции или фонограмм.

4.2 Наиболее полное понимание печатных документов достигается путем чтения всего текста. Такое полное чтение часто не реализуемо и не всегда является обязательным, одна ко индексатор должен гарантировать, что никакая полезная информация не упущена.

При обследовании документа индексатор должен обратить особое внимание на такие важные части текста:

- заглавие;
- реферат (если имеется);
- содержание;
- введение, начальные фразы глав и параграфов, заключение;
- иллюстрации, диаграммы, таблицы и сопроводительные надписи;
- слова или фразы, которые подчеркнуты или выделены некоторым другим способом (например напечатаны другим шрифтом).

Индексирование только на основе заглавия не рекомендуется. Реферат (если он есть) также не следует рассматривать как замену текста документа. Как заглавия, так и рефераты могут быть неадекватными тексту и не могут рассматриваться как надежный источник информации, необходимой индексатору.

4.3 В случае аудиовизуальной продукции или фонограмм приходится учитывать тот факт, что, например, аудиовизуальные, визуальные и звуковые носители требуют дополнительных технических процедур. На практике не всегда есть возможность обследования самого документа (например, прокручивания пленки). Индексирование в таком случае обычно выполняют на основе заглавия и/или краткого описания, хотя индексатор должен иметь возможность просмотреть треть или прослушать информацию непосредственно с носителя, если письменное описание оказывается неадекватным или неточным.

5 ИДЕНТИФИКАЦИЯ ПОНЯТИЙ

5.1 После обследования документа индексатор должен перейти к систематизированной идентификации тех понятий, которые являются существенными элементами предмета документа.

Заинтересованные организации должны установить контрольные списки тех параметров, которые считают важными в соответствующей предметной области. В перечисленных ниже вопросах иллюстрируются общие параметры, которые должны быть включены в такой контрольный список.

Описывается ли в документе объект деятельности?

Содержит ли предмет активное понятие (например действие, операция, процесс и т.п.)?

Идентифицирован ли объект деятельности?

Описывается ли в документе субъект действия?

Есть ли в документе указания на конкретные средства выполнения действия (например специальные инструменты, технологии или методы)?

Рассматривались ли эти параметры в контексте конкретного местоположения или среды?
Идентифицированы ли какие-нибудь зависимые или независимые переменные?

Рассматривался ли предмет в специальном аспекте, обычно не связанном с областью исследования (например социологическое исследование религии)?

Приведенные вопросы являются примерами общих параметров, которые применимы в любой предметной области. В каждой специальной дисциплине может потребоваться формулирование других вопросов.

5.2 Индексатор не должен обязательно идентифицировать все понятия, составляющие предмет документа. Отбор понятий зависит от цели индексирования.

Идентификация понятий может также зависеть от вида документа. Например, результаты индексирования на основе текстов книг, журнальных статей и т.п. чаще всего отличаются от результатов индексирования на основе рефератов или кратких обзоров.

Тщательность идентификации понятий связана с такими характеристиками индексирования как исчерпываемость и специфичность.

5.3 Исчерпываемость индексирования зависит от числа параметров (5.1), представленных понятиями, выявленными индексатором.

5.3.1 Индексатор должен быть в состоянии идентифицировать все понятия документа, которые имеют потенциальную значимость для пользователей той или иной информационной системы. Иногда две или более темы в пределах соответствующей предметной области рассматриваются независимо в одном документе. Они должны быть обработаны отдельно и, если потребуется, различными специалистами.

5.3.2 Границы соответствующей предметной области нельзя рассматривать слишком узко. Необходимо учитывать тот факт, что с развитием информационных сетей результаты индексирования, проведенного первоначально для одной группы пользователей (например ученых или технологов), могут использовать другие группы пользователей (например экономисты). С учетом этого индексаторам, например, научной и технической литературы, рекомендуется не игнорировать другие аспекты предмета, например, социальный и экономический.

5.3.3 При отборе понятий основным критерием всегда должна быть потенциальная значимость понятия как элемента выражения предмета документа и как параметра для поиска документа. При отборе понятий индексатор должен учитывать, насколько возможно, вопросы, поступающие в ту или иную информационную систему. Фактически, этот критерий отражает основную функцию индексирования. С учетом этого индексатор должен:

— выбрать понятия, наиболее подходящие с точки зрения некоторой заданной группы пользователей, помня при этом о задачах индексирования;

— изменить, в случае необходимости, процедуру индексирования, исходя из результатов обратной связи после обработки запросов. Однако такое изменение не должно противоречить структуре и логике языка индексирования.

5.3.4 Не устанавливаются никаких ограничений на количество терминов индексирования, а, значит, и понятий, приписанных одному документу. Количество понятий определяется только количеством информации, содержащейся в документе, с учетом задач индексирования. Установление каких-либо ограничений на количество понятий может привести к некоторой потере объективности в процессе индексирования и к потере информации, которая могла бы быть полезной при поиске. Если в рамках какой-то организации возникает потребность ограничить число терминов индексирования, отбор понятий следует определять суждениями индексатора относительно роли того или иного понятия в выражении предмета документа.

5.4 Специфичность индексирования определяется той степенью точности, с которой определенное понятие документа представляется средствами языка индексирования. Специфичность теряется, когда понятие представляется термином более общего значения.

Понятия следует идентифицировать с максимально возможной степенью специфичности. При определенных обстоятельствах более общие понятия могут оказаться более предпочтительными, например в таких случаях:

— когда высокая степень специфичности, по мнению индексатора, может отрицательно сказаться на характеристиках системы индексирования. Например, индексатор может решить заменить специфические наименования моделей оборудования более общими терминами, например, наименованием производителя или наименованием семейства моделей, особенно, если эти понятия находятся на границе соответствующей предметной области;

— если индексатор считает, что некоторая идея раскрыта не достаточно полно или указана автором только вскользь, и приписывает соответствующему понятию низкий вес.

6 ОТБОР ТЕРМИНОВ ИНДЕКСИРОВАНИЯ

6.1 При представлении выявленных понятий терминами индексирования индексатор должен придерживаться следующих правил (см. также приложение А):

— понятия, представленные в информационном языке, должны быть выражены соответствующими предпочтительными терминами;

— термины, выражающие новые понятия, следует проверять на точность и приемлемость с привлечением следующих средств: словарей и энциклопедий, признанных авторитетными в данной области; тезаурусов; классификационных систем.

Желательны также консультации со специалистами соответствующей предметной области, особенно с теми, кто имеет опыт в области индексирования.

6.2 Индексатор должен быть знаком с правилами применения этих средств индексирования. В частности, он должен представлять себе, что указанные средства могут накладывать определенные ограничения. Например, заданный заранее список предметных заголовков или состав разделов классификационной системы может не позволить точно выразить понятия, встреченные в документе. Если понятия выражены кодами классов, следует понимать, что эти коды обычно определяют более широкий или более узкий контекст (т.е. основной класс), который может и не совсем соответствовать обследуемому документу.

6.3 Если информационный язык включает тезаурус, количество терминов, приписываемых документу, может быть уменьшено без потерь, т. к. родовые и другие априори существующие отношения могут быть установлены непосредственно на основе тезауруса. В случае использования тезауруса для представления определенного понятия следует выбирать самый специальный термин.

6.4 В некоторых системах индексирования используют индикаторы роли, связи, веса и т.п. Индексатор должен быть знаком со всеми специальными правилами, связанными с этими механизмами.

6.5 На практике индексатор зачастую встречает понятия, отсутствующие в существующем тезаурусе или классификационной системе. В зависимости от используемой системы эти понятия могут быть обработаны различным образом, например:

— они могут быть выражены терминами или дескрипторами, которые сразу же вводятся в информационный язык;

— они могут быть временно представлены более общими терминами, а новые понятия могут быть предложены как потенциальные для более позднего пополнения информационного языка.

7 КОНТРОЛЬ КАЧЕСТВА ИНДЕКСИРОВАНИЯ

7.1 Качество индексирования зависит от следующих факторов:

— квалификации и опыта индексатора;

— качества средств индексирования.

Качество индексирования в значительной степени определяется непротиворечивостью индексирования, состоящей в том, что состав терминов индексирования, приписанных некоторому документу, и достигнутые при индексировании уровни исчерпываемости и специфичности являются в идеале независимыми от индексатора, и остаются неизменными на протя-

жении функционирования конкретной системы индексирования. На практике не всегда возможно достичь идеала непротиворечивости, но стремление к нему является важным фактором, влияющим на характеристики системы индексирования, особенно в случае обмена информацией между различными организациями в сети.

7.2 Обязательным условием достижения непротиворечивости индексирования является полное беспристрастие со стороны индексатора при идентификации понятий и выборе терминов индексирования. В случае, когда есть большая группа индексаторов, работающих в разных местах, например в децентрализованных системах, для достижения непротиворечивости индексирования рекомендуется введение этапа централизованной проверки с обратной связью с индексаторами.

7.3 Для обеспечения достаточного качества индексирования индексатор должен обладать соответствующими знаниями в той предметной области, к которой относят обследуемые документы. Он должен понимать термины, встречающиеся в документах, а также правила применения конкретного информационного языка.

В случае индексирования документов на иностранных языках индексатор должен владеть соответствующим языком или иметь возможность обратиться за помощью к соответствующим специалистам.

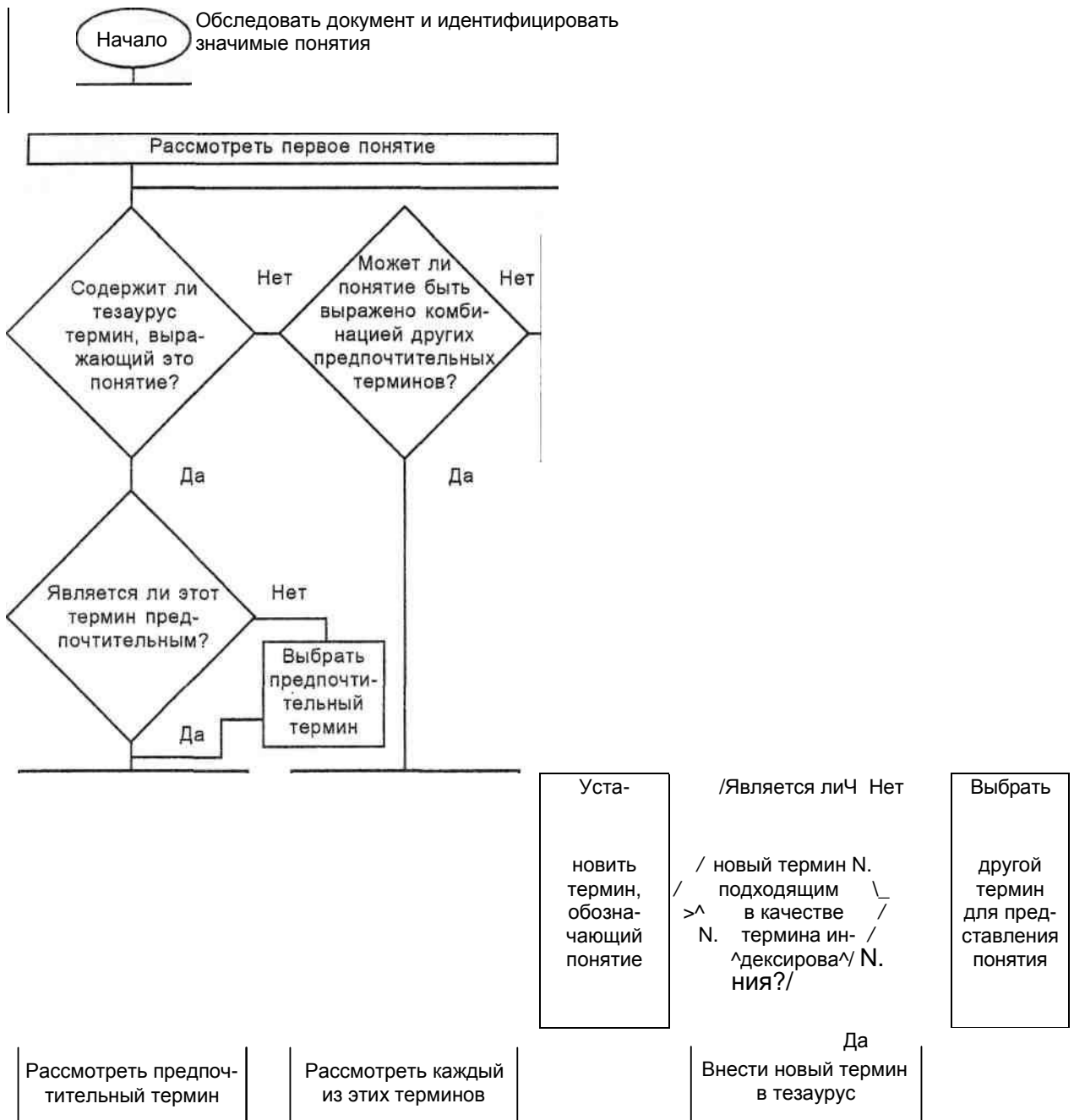
7.4 Повышению качества индексирования будут способствовать контакты индексаторов с пользователями. Например, с целью выявления ложных терминологических комбинаций, которые могут порождать нерелевантную выдачу при информационном поиске.

7.5 Качество индексирования зависит также от гибкости используемого информационного языка. Желательно, чтобы он допускал включение новых терминов и внесение изменений в терминологию.

7.6 Качество индексирования следует проверять по возможности путем анализа результатов информационного поиска, например с использованием коэффициентов полноты и точности.

ПРИЛОЖЕНИЕ А
(справочное)

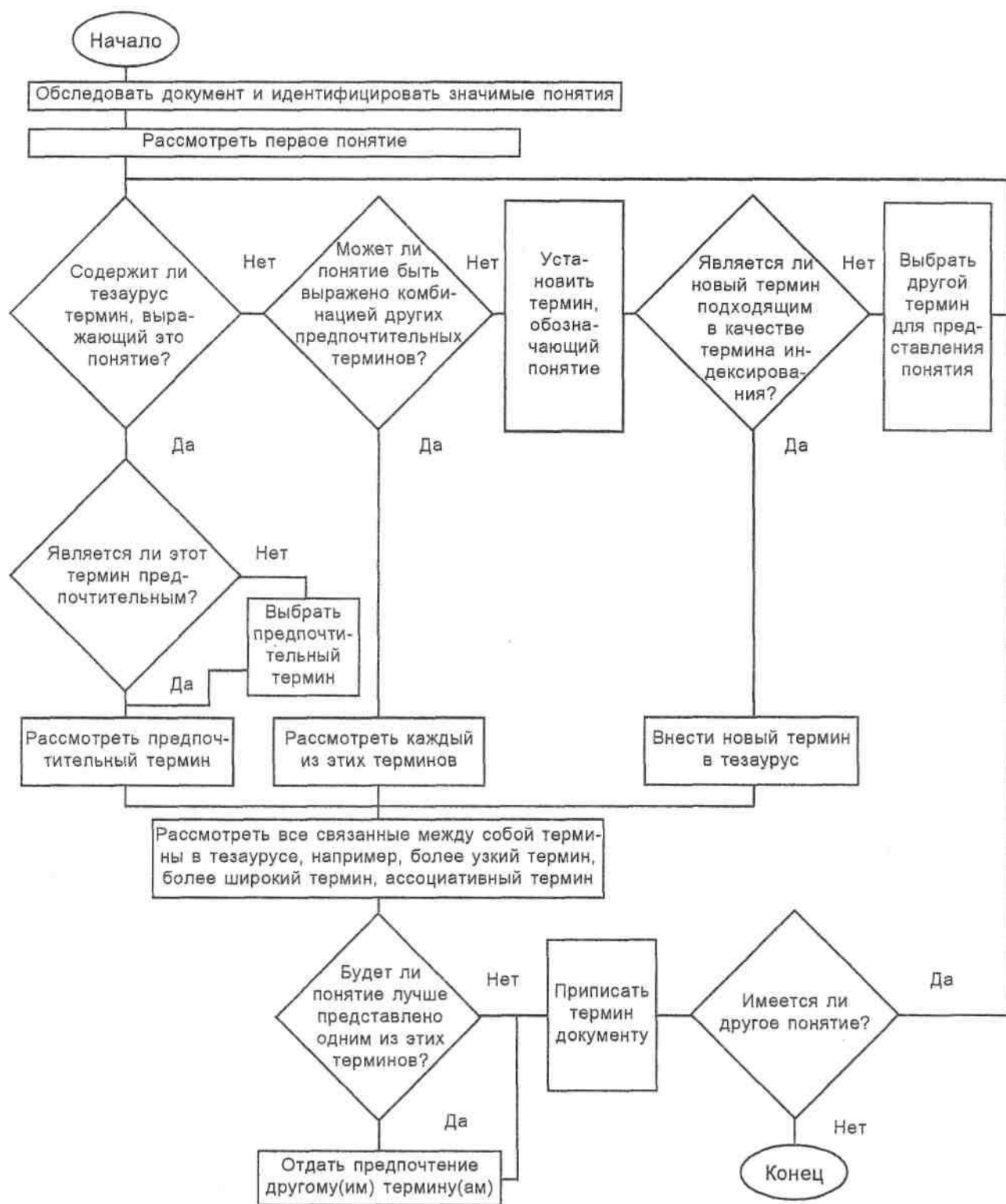
БЛОК-СХЕМА ПРОЦЕДУРЫ ИНДЕКСИРОВАНИЯ
С ИСПОЛЬЗОВАНИЕМ ТЕЗАУРУСА



Рассмотреть все связанные между собой термины в тезаурусе, например, более узкий термин, более широкий термин, ассоциативный термин



Нет
Отдать предпочтение
другому(им) термину(ам)



ГОСТ 30671-99

УДК 002:001.815:006.354

01.140.20

T62

Ключевые слова: общая методика, информация, документация, индексирование, предмет документа, термин индексирования, информационный язык, тезаурус, классификационная система, информационная деятельность.

Редактор **Г. Ярмиш**
Техжчний редактор **Т. Новікова**
Коректор **Т. Нагорна**
Комп'ютерна верстка **І. Сохач**

Шдписано до друку 02.02.2001. Формат 60 x 84 1/8. Ум.
друк. арк. 2,79. Зам. 3 11 Ц1на догов1рна.

Вщдт пол1графи науково-техжчних видань УкрНД1СС1
03150, КиГв-150, вул. Горького, 174