

кому та Мадридському критеріям набуття членства в Європейському Союзі, якій включає адаптацію законодавства, утворення відповідних інституцій та інші додаткові заходи, необхідні для ефективного правотворення та правозастосування.

Як визначається у проекті, метою адаптації законодавства України до законодавства ЄС є досягнення відповідності правової системи України *acquis communautaire* з урахуванням критеріїв, що висуваються Європейським Союзом до держав, які мають намір вступити до нього.

Адаптація законодавства є пріоритетною складовою процесу інтеграції України до ЄС.

Успішна інтеграція українського законодавства до стандартів ЄС неможлива без вжиття ряду попередніх заходів, насамперед щодо адаптації термінологічної бази. Зазначене питання має безпосереднє відношення до лінгвістичного забезпечення, тобто узгодження українських термінів із термінами стандарту Європейського Союзу – інформаційно-пошукового тезауруса EUROVOC.



С.ДОРОГИХ, завідувач лабораторії НДЦ правової інформатики АПрН України

УКРАЇНСЬКА ВЕРСІЯ ІНФОРМАЦІЙНО-ПОШУКОВОГО ТЕЗАУРУСА EUROVOC

Історія створення інформаційно-пошукового тезауруса EUROVOC

Робота зі створення тезауруса почалася у 1982 р., коли перед групою фахівців з документалістів та бібліотечних працівників бібліотеки Європейського Парламенту, Відділу офіційних публікацій ЄС та інших установ було поставлено завдання створити багатомовний тезаурус, який охоплював би галузі діяльності інститутів ЄС. Проект був заснований на базі існуючого одномовного тезауруса, створеного Європейським Парламентом на основі особливостей змісту макротезауруса OECD (Організації економічного співробітництва і розвитку), початкової версії 1978 р. тезауруса ECLAS та вільних дескрипторів, за якими відбувалася їх індексація різними структурними підрозділами ЄС.

У 1984 р. Європейський Парламент і Відділ офіційних публікацій видали перший випуск тезауруса EUROVOC у двох томах (алфавітний тезаурус і предметний тезаурус) сімома мовами (датською, голландською, англійською, французькою, німецькою, грецькою, італійською).

Друге видання тезауруса, опубліковане в 1987 р., стало результатом необхідних змін першого видання. Потреба в змінах була викликана індексацією в Європейському Парламенті приблизно 25 тис. документів, а також розробкою Офіційного Журналу у Відділі офіційних публікацій.

До семи мов додали ще дві – іспанську та португальську, а до перших двох представлень (алфавітного та тематичного) – багатомовне та пермутаційне представлення і термінограми.

Робота з внесення змін і доповнень до тезауруса проводилася бібліотекарями і документалістами Європейського Парламенту та представниками Відділу публікацій, які займаються індексацією. Мовні еквіваленти були підібрані перекладачами і перевірені Відділом термінології Європейського Парламенту.

Друге видання складалося з 21 розділу, 126 мікротезаурусів, 5359 дескрипторів. Під час його створення було додано 1004 нових дескрипторів і 936 видалено.

Враховуючи постійно зростаючий інтерес до ІПТ EUROVOC, Європейський Парламент і Відділ офіційних публікацій, яким належить ІПТ EUROVOC, у 1989 р. вмірили організувати перший семінар для користувачів, на якому ухвалили рішення про управління і зміст тезауруса.

Після набуття чинності Угоди про Європейський Союз і в результаті політичних змін, що відбулися в Європі, було прийнято рішення про створення нової версії ІПТ EUROVOC. З урахуванням того, що найбільш вживаною була французька версія тезауруса, були переглянуті та відредаговані переклади іншими мовами.

Третє видання вийшло в 1995 р. У його складі – 21 розділ, 127 мікротезаурусів, 5933 дескриптори (при його підготовці до другого видання було додано 566 нових дескрипторів та 1 мікротезаурус; 26 дескрипторів було вилучено).

На сьогодні вже діє четверте видання ІІТ EUROVOC. Воно має три форми представлення: абетково-пермутаційну; тематичну; багатомовну.

Загальні відомості про інформаційно-пошукові тезауруси

Накопичені великі фонди і потоки інформації, в яких необхідно знаходити дані, що відповідають запитам користувача, потребують створення ефективного інструментарію для обробки, пошуку та аналізу. Одним з таких інструментів є інформаційно-пошукові тезауруси (ІІТ). Вони були створені для індексування документів та виконання машинного пошуку. Визначення предметної суті документів, вибір термінів, сукупність яких стисло передає зміст, зазначення відношень між поняттями, представленими цими термінами, основні завдання, що вирішуються при створенні тезауруса¹.

Традиційно при написанні документів автори вільно використовують засоби природної мови. Водночас, людина, яка бажає знайти певний документ, може зіткнутися з труднощами, пов'язаними з тим, що:

- одне й те саме поняття може бути виражено рядом синонімів або понять, які частково збігаються (наприклад, квартира, помешкання, житлове приміщення тощо). При визначенні ключових слів або при контекстному пошуку документ позначається тільки одним з перелічених вище синонімів. Відповідно, користувач, який виконує пошук за одним із синонімів, не знайде документи, котрі відповідають його запиту, але в тексті яких те саме поняття виражене іншими словами. При цьому пошук документів на мові, відмінній від мови пошуковця, дуже ускладнюється;

- один і той же термін може мати різні значення (наприклад, слово “камера” може вживатися у значенні пристрою для відеозйомки або тюремного приміщення). В результаті використання такого слова при пошуку, будуть знайдені документи, які не мають відношення до теми пошуку. Тезаурус виключає багатозначність.

Нарешті, тезауруси надають можливість користувачеві отримати загальну інформацію про термінологічний апарат та структуру предметної області, що його цікавить, отримати відомості про зв'язки з суміжними областями.

Наявність тезауруса як такого не може допомогти користувачеві знайти необхідний документ. Потрібно провести роботу індексування масиву інформації, у якому відбуватиметься пошук.

Тобто тезаурус є предметним покажчиком, який описує певну предметну область за допомогою спеціально визначених слів (такі слова мають назву “дескриптори”), встановлює відношення дескрипторів до їх синонімів або близьких понять, котрі прирівнюються до синонімів (аскрипторів*), визначає внутрішню структуру предметної області шляхом встановлення відношень між різними дескрипторами, формалізує природну мову та є інструментом для індексації та пошуку документів.

Слід також зазначити, що з часом предметна область, яку описує тезаурус, буде змінюватись, і, відповідно, виникне необхідність внесення змін до нього.

Багатомовні тезауруси** розширюють функції одномовних і полегшують міжнародний обмін документами. Вони будуються за принципом тотожності структури. Тобто для кожної з мов, яку підтримує багатомовний тезаурус, існує однакова кількість дескрипторів, котрі мають між

¹ДСТУ 4032-20001 (ISO 2788:1986) Одномовний тезаурус. Методика розроблення. – К., 2001.

* термін тезауруса, який не повинен використовуватися для вираження поняття і має посилання на один чи декілька дескрипторів, що використовуються замість нього.

** тезауруси які містять дескриптори, як правило, недескриптори, взяті з кількох природних мов, що виражають еквівалентні поняття в кожній з цих мов.

собою однакові семантичні зв'язки. Водночас, система аскрипторів розробляється окремо для кожної мови і відображає її специфіку.

Дескриптори з різних мов, що стосуються одного й того самого поняття, зазвичай, трактуються як точні еквіваленти. Проте в окремих випадках така еквівалентність є неточною або частковою. При неточній еквівалентності поняття, що виражені дескрипторами, трактуються як такі, що позначають ті самі множини об'єктів та явищ, але склад цих множин дещо варіюється. При частковій еквівалентності дескриптори відображають одне й те саме поняття, проте в якійсь з мов це поняття визначається ширше або вужче². Ця проблема вирішується широким використанням аскрипторів та приміток.

Загалом, багатомовні дескриптори надають можливість виконати пошук мовою користувача та отримати документи іншою мовою (див. Рис. 1).



Рис. 1

Призначення інформаційно-пошукового тезауруса EUROVOC

ІПТ EUROVOC виконує всі функції, покладені на багатомовні тезауруси, і відповідає міжнародним стандартам ISO 2788-1986 “Керівні принципи створення і розробки одномовних тезаурусів”³ та ISO 5964-1985 “Керівні принципи створення і розробки багатомовних тезаурусів”⁴.

При розробці вітчизняного правового поля постає питання гармонізації українського законодавства із міжнародними правовими нормами, а також його презентації іноземним користувачам, які бажають ознайомитися з нормативно-правовими актами України. У Національній програмі інформатизації інтеграція України у світовий інформаційний простір визначається як одне з основних завдань⁵. В той же час майже не існує інструментарію для реалізації цих завдань на практиці. За цих умов ІПТ EUROVOC може стати одним з таких інструментів.

Багатомовний тезаурус EUROVOC був спеціально розроблений для роботи з документальною інформацією, якою володіють інститути Європейського Союзу.

Тезаурус EUROVOC використовується бібліотеками інститутів Європейського Союзу, документальними службами, документальними базами даних і користувачами відповідних служб, а саме:

- Європейським Парламентом разом з його документальною базою даних EPOQUE, в якій містяться матеріали парламентських документів і бібліотечний каталог;
- Відділом офіційних публікацій ЄС з його базами даних (особливо CATEL, каталогом публікацій) і Офіційним журналом ЄС, які прийняли EUROVOC для своєї документації і, таким чином, мають доступ до існуючої європейської документації, яка вже індексована на базі того ж тезауруса охоплює всі офіційні публікації інститутів ЄС, документації Європейського Парламенту і

² ДСТУ 4031-20001 (ISO 5964:1985) Багатомовний тезаурус. Методика розроблення. – К., 2001.

³ ISO 2788-1986 - Guidelines for the establishment and development of monolingual thesauri.

⁴ ISO 5964-1985 - Guidelines for the establishment and development of multilingual thesauri.

⁵ Закон України “Про Національну програму інформатизації” від 04.02.1998 № 74/98-ВР.

змісту Офіційного журналу ЄС. EUROVOC використовується парламентами цілого ряду країн (наприклад, Бельгійською Палатою Представників, Іспанським Конгресом, Асамблеєю Португальської Республіки тощо).

У 2001 році видавництво Державної Думи видало російську версію тезауруса EUROVOC.

EUROVOC виступає в ролі моста, який з'єднує всі системи документації, яка має відношення до діяльності Європейського Союзу, як всередині інститутів ЄС на національному та регіональному рівнях, так і в приватному секторі. Будь-яка посадова чи приватна особа може використовувати тезаурус EUROVOC, необхідно лише мати його носії⁶.

При роботі з тезаурусом EUROVOC потрібно враховувати як переваги, так і недоліки (обмеження) цього тезауруса, викликані його специфікою та метою створення.

До основних переваг відносяться:

- термінологічна стандартизація словників введена до документальної бази даних для запитів у цій базі, що робить пошук досить ефективним;
- завдяки багатомовності відбувається пошук документів, написаних різними мовами, хоча запит здійснюється мовою користувача бази даних;
- простота використання EUROVOC в якості мови індексування дозволяє різним інформаційним службам працювати в єдиній системі, що зменшує витрати на обробку документів.

Проте існує й цілий ряд недоліків:

- EUROVOC був розроблений так, щоб відповідати вимогам загальних систем документації, які охоплюють діяльність ЄС, але він не підходить для спеціальної, вузькогалузевої документації;
- EUROVOC орієнтовано на діяльність інститутів ЄС, що призводить до обмеження обробки документальної інформації на національному та регіональному рівнях.

Сфери практичного застосування тезауруса EUROVOC в Україні

Адаптація законодавства України до законодавства ЄС. Орієнтація України на співпрацю та інтеграцію з Європейським Співтовариством визначена як один з основних напрямів державної політики України та затверджена у багатьох нормативно-правових документах. Так, у 1998 році було видано Указ Президента “Про затвердження Стратегії інтеграції України до Європейського Союзу”⁷, в якому одним з основних напрямів визначена адаптація законодавства України до законодавства ЄС. Згідно з цим Указом адаптація законодавства України до законодавства ЄС полягає у зближенні із сучасною європейською системою права, що забезпечить розвиток політичної, підприємницької, соціальної, культурної активності громадян України, економічний розвиток держави в рамках ЄС і сприятиме поступовому зростанню добробуту громадян до рівня, що склався в державах-членах ЄС. Адаптація законодавства України передбачає реформування її правової системи та поступове приведення у відповідність із європейськими стандартами охоплює приватне, митне, трудове, фінансове, податкове законодавство, законодавство про інтелектуальну власність, охорону праці, охорону життя та здоров'я, навколишнє природне середовище, захист прав споживачів, технічні правила і стандарти, транспорт, а також інші галузі, визначені Угодою про партнерство та співробітництво.

Для успішної інтеграції українського законодавства до стандартів ЄС згідно з Постановою Кабінету Міністрів “Про Концепцію адаптації законодавства України до законодавства Європейського Союзу”⁸ передбачається інформаційне забезпечення, яке включає:

⁶ Тезаурус “Eurovoc”: Посібник для користувачів. – 1992.

⁷ Указ Президента “Про затвердження Стратегії інтеграції України до Європейського Союзу” від 11.06.1998 № 615/98.

⁸ Постанова Кабінету Міністрів “Про Концепцію адаптації законодавства України до законодавства Європейського Союзу” від 16.08.1999 № 1496.

- постійне інформування органів державної влади про чинне законодавство ЄС, зміни і доповнення до нього, забезпечення ефективного та повного доступу до європейських правових, бібліографічних та інших баз даних учасників процесу адаптації законодавства;

- організацію роботи з перекладу актів законодавства ЄС та надання консультацій з питань використання юридичної термінології, забезпечення діяльності Центру європейського та порівняльного права;

- організацію роботи з наукового дослідження процесів уніфікації та гармонізації права в державах - членах ЄС, державах Центральної і Східної Європи, що є кандидатами у члени ЄС.

Адаптація ІПТ EUROVOC в Україні створює передумови для виконання вирішення вищеперерахованих завдань інформаційного забезпечення інтеграції українського законодавства. Так, постійне інформування органів державної влади про чинне законодавство ЄС і доступ до баз даних потребує використання відповідного пошукового інструментарію, яким у країнах ЄС є EUROVOC. Пошукові можливості ІПТ EUROVOC дозволяють відібрати нормативно-правові документи країн ЄС, котрі містяться у правових базах даних, які відповідають певній темі, що сформульована у пошуковому запиті. Завдяки тому, що використання ІПТ EUROVOC дозволяє виконувати пошук мовою користувача (для цього EUROVOC повинен бути адаптований до відповідної мови) і знайти документи викладені будь-якою з мов країн ЄС, на яку перекладено EUROVOC (на сьогодні EUROVOC офіційно перекладено одинадцятьма мовами), український законотворець може відібрати необхідні йому документи і визначити документи та напрямки для першочергового перекладу. Крім того, розподіл документів за темами створює передумови для аналізу законодавства країн ЄС, дозволяє порівнювати співвідношення та зміни у законодавстві за певними галузями в різних країнах ЄС.

Міжнародний обмін інформацією. Для вивчення іноземного законодавства та презентації українського законодавства громадянам інших країн необхідно мати ефективну систему міжнародного обміну інформацією. Завдяки ІПТ EUROVOC створюється багатомовний словник, за яким відбувається індексування нормативно-правових документів як у країнах ЄС, так і в країнах, що не є його членами, але адаптують ІПТ EUROVOC для власних потреб. Єдина система індексування дозволяє формувати єдині інформаційні запити до правових баз даних будь-якої з країн, яка підтримує ІПТ EUROVOC. Нагадаємо також, що при використанні ІПТ EUROVOC запит можна формувати мовою користувача, а отримувати документи мовою оригіналу.

Індексування нормативної бази. Окрім організації пошуку вітчизняного користувача в іноземних базах, ІПТ EUROVOC може з певними обмеженнями виступати в якості інструменту для індексації українського законодавства. Використання ІПТ EUROVOC дає можливість одночасно розпочати вирішення проблем відсутності в Україні єдиного класифікатора нормативно-правових документів, аналізу та презентації українського законодавства для іноземного користувача.

Пошукова система. Використовуючи ІПТ EUROVOC в якості інструмента для пошуку, ми маємо розглядати його в двох аспектах: для пошуку в іноземних базах даних та для пошуку у вітчизняному законодавстві.

ІПТ EUROVOC використовується для індексування та пошуку у базах даних EUR-Lex, CELEX та EPOQUE, електронному каталозі CATEL тощо.

База даних EUR-Lex – це портал законодавства Європейського Союзу. Він поділяється на секції:

- Офіційний журнал ЄС (Official Journal) видається 5 разів на тиждень одинадцятьма мовами, складається з 2 серій – L (всі нормативні акти), C (інформація, підготовчі роботи, нотатки та рекомендації) та додаткову S (пропозиції), що публікується в базі даних TED;

- угоди;

- законодавство (у цій секції зберігаються всі законодавчі акти, що були прийняті ЄС відповідно до угод);

- підготовчі документи (у цій секції можна ознайомитись з проектами документів на всіх стадіях законотворчого чи бюджетного процесу);

- прецедентне право (містить рішення та ухвали Європейського суду, суду першої інстанції та вердикти генерального адвоката);
- парламентські запити (ця секція присвячена усним та письмовим запитам членів Європарламенту та відповідям на них);
- документи, що становлять загальний інтерес (документи, які присвячені загальнолюдським, загальноєвропейським питанням, такі як Білі сторінки, інформація та зв'язок, Зелені сторінки тощо).

Ще один з ресурсів, який надається на сайті ЄС – це анотації до законодавства (Summaries of legislation). База даних цього ресурсу містить детальну інформацію одинадцятьма офіційними мовами щодо всіх видів діяльності Європейського Союзу. Вона містить короткі виклади і посилання на публікації заходів, що проводяться Європейським Союзом в усіх сферах діяльності, включно з попередньою роботою і відповідними звітами. Сайт регулярно оновлюється відповідно до законодавчих процедур. Короткі виклади законодавства також містять зв'язки із повними текстами рішень, правил, директив.

Pre-Lex - база даних з законотворчої процедури. Вона надає можливість простежити за головними станами процесу прийняття рішення Комісією та іншими установами і забезпечує інформацію про:

- етапи прийняття законодавчого акта;
- рішення, прийняті кожною установою на всіх стадіях проходження документа;
- список осіб, залучених до процесу прийняття документа (члени Комісій, доповідачі та ін.);
- довідкові документи.

База даних Pre-Lex надає можливість прослідкувати за поданням пропозицій та роботою різних установ, що беруть участь у процесі підготовки документа (Європейський Парламент, Рада, Економічний і Соціальний Комітет, Комітет Регіонів і т. д.) від передачі пропозиції до Ради або Європарламенту до прийняття або відхилення проекту.

За допомогою гіперпосилань з бази Pre-Lex можна отримати прямий доступ до електронних текстів (Офіційний Журнал, документи Європейського Парламенту, преса Ради та ін.).

База даних CELEX містить договори про створення та основні засади роботи Європейського Союзу; міжнародні договори та інші документи, пов'язані із зовнішньою політикою ЄС; обов'язкові та необов'язкові правові акти; судову практику; парламентські запити тощо.

База даних EPOQUE⁹ містить інформацію з усіх видів діяльності Європейського парламенту, яка поділяється на:

- документи Європейського парламенту;
- відомості щодо законодавчої та парламентської процедур (в цих відомостях у документальному вигляді знаходять своє відображення всі етапи процесу прийняття рішень).

База даних CATEL¹⁰ містить електронні каталоги видань і документів ЄС, а також документів, опублікованих в Офіційному журналі ЄС.

З базою OEIL ви можете відслідковувати процедуру прийняття рішення, з'ясувати, якої стадії дійшов процес проходження документа та скориставшись розкладом, визначити термін наступної стадії. База даних дає змогу скласти думку про роботу Парламенту і його комітетів, щорічну робочу програму Комісії та щодо пропозицій президентів Ради.

Також можна виконувати пошук документів країн ЄС або країн, що адаптували ІПТ EUROVOC для індексації власного законодавства.

Відмітимо, що доступ до таких баз даних можна отримати через Інтернет. У поданій нижче Таблиці вказані адреси в Інтернет деяких з цих баз.

При індексації вітчизняного законодавства за допомогою ІПТ EUROVOC створюється пошуковий масив, за яким можна виконувати пошук серед нормативно-правових документів України. Враховуючи, що при індексації документів використовуються дескриптори, які визначають найголовніше у

⁹ EPOQUE: European Parliament On-line QUERy system. User manual – Guided Search Mode / European Parliament. Directorate General for Research. – Luxembourg, 1995.

¹⁰ CATEL: Electronic catalogue. User manual / Office for Official Publications of the European Communities. – Luxembourg, 1992.

змісті документа та відсікають “інформаційний шум” (документи, в яких слово, за яким виконувався пошук, використане випадково і документ по суті не відповідає запиту користувача). Крім того, завдяки тому, що дескриптори згруповані за темами і між ними встановлені зв’язки, можна виконувати пошук за суміжними темами, отримати список близьких понять, що характеризують галузь і за якими можна виконати пошук. В результаті використовуючи ІПТ EUROVOC ми отримуємо ефективну пошукову систему не тільки в іноземному, але й у вітчизняному законодавстві.

Таблиця

Назва бази	Адреса в Інтернет	Примітки
EUR-Lex	http://europa.eu.int/documents/eur-lex/index_en.htm	мови ЄС*
Офіційний журнал ЄС	http://europa.eu.int/eur-lex/en/search/search_oj.html	мови ЄС
Анотації до законодавства ЄС	http://europa.eu.int/scadplus/scad_en.htm	мови ЄС
Pre-Lex	http://europa.eu.int/documents/eur-lex/index_en.htm	мови ЄС
CELEX	http://europa.eu.int/documents/celex/index_en.htm	доступ до бази платний; мови ЄС
OEIL	http://wwwdb.europarl.eu.int/dors/oeil/en/default.htm	англ. та фр.
TED	http://ted.publications.eu.int/static/home/en/homepage.ini	мови ЄС

Аналіз українського законодавства. Структура, лексичний склад, наповненість документами різних тем можуть бути використані для різних аспектів співставлення законодавства України до норм, що закріплені у європейському законодавстві. А це є одним з напрямів діяльності по адаптації законодавства України до законодавства ЄС, що визначається як один з пріоритетних напрямків розбудови України.

Аналіз термінології, затвердженої в українському законодавстві, на предмет відповідності термінології законодавства країн ЄС. Тезаурус EUROVOC містить понад 6000 тисяч еквівалентів понять одинадцятьма мовами (англійською, голландською, грецькою, датською, іспанською, італійською, німецькою, португальською, фінською, французькою та шведською), їх зв’язки та різні способи вираження для кожної з мов. Таким чином, ІПТ EUROVOC сам є довідковою понятійно-термінологічною базою даних. Робота над адаптацією ІПТ EUROVOC для роботи в Україні передбачає співставлення термінологічного апарату ІПТ EUROVOC з масивом термінів, які мають своє визначення у нормативних актах України.

Кількісні показники інформаційно-пошукового тезауруса EUROVOC

Четверта версія тезауруса EUROVOC містить незалежно від мови (маються на увазі офіційні переклади. Деякі країни, що не є членами ЄС, наприклад Росія, під час адаптації ІПТ EUROVOC для власного використання пішли шляхом додавання власних дескрипторів): 21 розділ; 127 мікротезаурус; 6439 дескрипторів (з них 511 верхнього рівня); 6448 ієрархічних зв’язків; 3501 асоціативних зв’язків¹¹.

Кількість аскрипторів та приміток різна для кожної з мовних версій і коливається, наприклад, від 5211 (фінська версія) до 9210 (італійська версія). Російська версія ІПТ EUROVOC за станом на 1 липня 2000 р. містить: 9648 дескрипторів (серед яких 545 верхнього рівня); 1393 ас-

* Усі посилання в таблиці подаються на версію з англійським інтерфейсом, з якої можна перейти на інтерфейс на будь-якій з 11 інших офіційних мов ЄС.

¹¹ Thesaurus EUROVOC: Vol. 1-3 / European Communities. – Luxembourg: Office for Official Publications of the European Communities, 2000. – Ed. 4. – English Language.

крипторів; 9138 ієрархічних зв'язків; 5681 асоціативних зв'язків; 1124 зв'язків від аскриптора до дескриптора, що його замінює; 175 зв'язків від аскриптора до комбінації дескрипторів, що його замінює; 267 зв'язків від аскриптора до групи дескрипторів, яка альтернативно його замінює; 2898 приміток¹².

На першому етапі при розробці української версії ІПТ EUROVOC акцент робиться на відповідності української версії офіційним версіям країн ЄС. У подальшому для подолання обмежень щодо використання ІПТ EUROVOC при індексації національного законодавства планується широкі залучення аскрипторів та приміток.

Офіційні мовні версії інформаційно-пошукового тезаурусу EUROVOC

На сьогодні ІПТ EUROVOC офіційно видано одинадцятьма мовами країн ЄС: англійською; голландською; грецькою; датською; іспанською; італійською; німецькою; португальською; фінською; французькою; шведською.

Ця кількість постійно зростає. Так, перша версія ІПТ EUROVOC була представлена 7 мовами, друга та третя версія – 9, а четверта – 11 мовами. Безперечно, що зі збільшенням країн-членів ЄС кількість мов, на яких здійснено офіційний переклад, буде збільшуватися.

Усі вищеперераховані мови мають однаковий статус, тобто кожний дескриптор має еквіваленти в усіх інших мовних версіях. Типи зв'язків між дескрипторами теж збігаються (див. Рис. 2).



Рис. 2

Відповідно, в усіх мовних версіях однакова кількість розділів, мікротезаурусів, дескрипторів, зв'язків між дескрипторами.

Тематичні розділи інформаційно-пошукового тезаурусу EUROVOC

У тезаурусі EUROVOC виділені основні розділи, пов'язані з різними видами діяльності органів, установ та інститутів ЄС:

ПОЛІТИКА
МІЖНАРОДНІ ВІДНОСИНИ
ЄВРОПЕЙСЬКІ СПІВТОВАРИСТВА
ПРАВО
ЕКОНОМІКА
ТОРГІВЛЯ
ФІНАНСИ
СОЦІАЛЬНІ ПИТАННЯ
ОСВІТА, КОМУНІКАЦІЯ
НАУКА

¹² Москаленко Т.А. Предисловие // Информационно-поисковый тезаурус. Русская версия тезауруса EUROVOC: Т. 1-3. – М.: Издание Государственной Думы, 2001.

ПІДПРИЄМНИЦТВО, КОНКУРЕНЦІЯ
 ПРАЦЯ, ЗАЙНЯТІСТЬ
 ТРАНСПОРТ
 НАВКОЛИШНЄ СЕРЕДОВИЩЕ
 СІЛЬСЬКЕ ГОСПОДАРСТВО, ЛІСОВЕ ГОСПОДАРСТВО, РИБНЕ ГОСПОДАРСТВО
 СІЛЬСЬКОГОСПОДАРСЬКІ ПРОДУКТИ ХАРЧУВАННЯ
 ВИРОБНИЦТВО, ТЕХНОЛОГІЇ, ДОСЛІДЖЕННЯ
 ЕНЕРГЕТИКА
 ПРОМИСЛОВІСТЬ
 ГЕОГРАФІЯ
 МІЖНАРОДНІ ОРГАНІЗАЦІЇ.

Розподіл дескрипторів за тематичними розділами, як відмічають розробники EUROVOC, певною мірою довільний. Багато дескрипторів за своїм змістом можуть відноситися не до одного, а до декількох мікротезаурусів. Проте, для того, щоб уникнути полієрархії (наявності у дескриптора декількох ширших дескрипторів), яка б збільшила обсяг тезауруса та ускладнила б користування ним, розробники визначили, що один дескриптор знаходиться тільки в одному предметно-тематичному мікротезаурусі. Подібне рішення не викликає нарікань ні у користувачів, які виконують пошук за допомогою EUROVOC, ні у індексаторів, тому що як при побудові запитів, так і при індексуванні документів, можна робити комбінації з дескрипторів, які належать до різних розділів та мікротезаурусів.

Виключення з цього правила допускається лише стосовно розділів “ГЕОГРАФІЯ” та “МІЖНАРОДНІ ОРГАНІЗАЦІЇ”.

Так, наприклад, дескриптор “ІРАН” має ширший дескриптор “КРАЇНИ ПЕРСЬКОЇ ЗАТОКИ” у мікротезаурусі “АЗІЯ ТА ОКЕАНІЯ”, а також ширший дескриптор “КРАЇНИ ОПЕК” у мікротезаурусі “ЕКОНОМІЧНА ГЕОГРАФІЯ”.

Структура подання матеріалу в українській друкованій версії ІПТ EUROVOC

ІПТ EUROVOC побудовано за принципом від загального до часткового. На верхньому рівні знаходяться розділи та мікротезауруси, які складають загальну предметно-класифікаційну систему (див. Рис. 3).

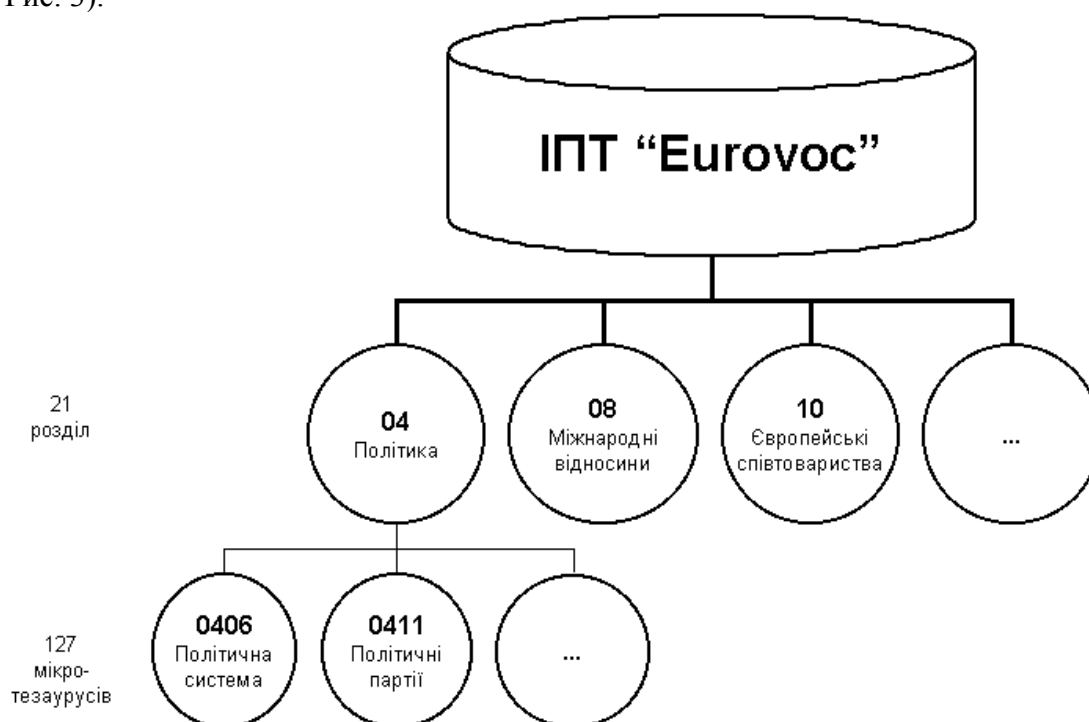


Рис. 3

У тезаурусі EUROVOC виділено 21 розділ, які ідентифікуються двозначним кодом та назвою на природній мові.

Приклад: 04 ПОЛІТИКА.

Кожен з розділів поділено на тематичні мікротезауруси (загалом їх 127). Мікротезауруси ідентифікуються чотирьохзначним кодом (перші дві цифри співпадають з кодом розділу, до складу якого входить мікротезаурус) та назвою на природній мові.

Приклад: 0406 ПОЛІТИЧНА СИСТЕМА.

Враховуючи можливі зміни в EUROVOC, розробники пронумерували розділи та мікротезауруси не підряд, а залишаючи проміжки (04, 08, 10, 12, 16, 20, ...). Числові коди розділів та мікротезаурусів співпадають в усіх мовних версіях.

На нижньому рівні знаходяться дескриптори, які складають лексичну частину тезауруса (див. Рис. 4).

Тут структура тезауруса визначається семантичними відношеннями, до яких належать:

– **примітки** (довідкова інформація або інформація стосовно використання дескриптора)

Приклад: **КРАЇНИ ОПЕК**

прим.: Країни-члени організації країн-експортерів нафти

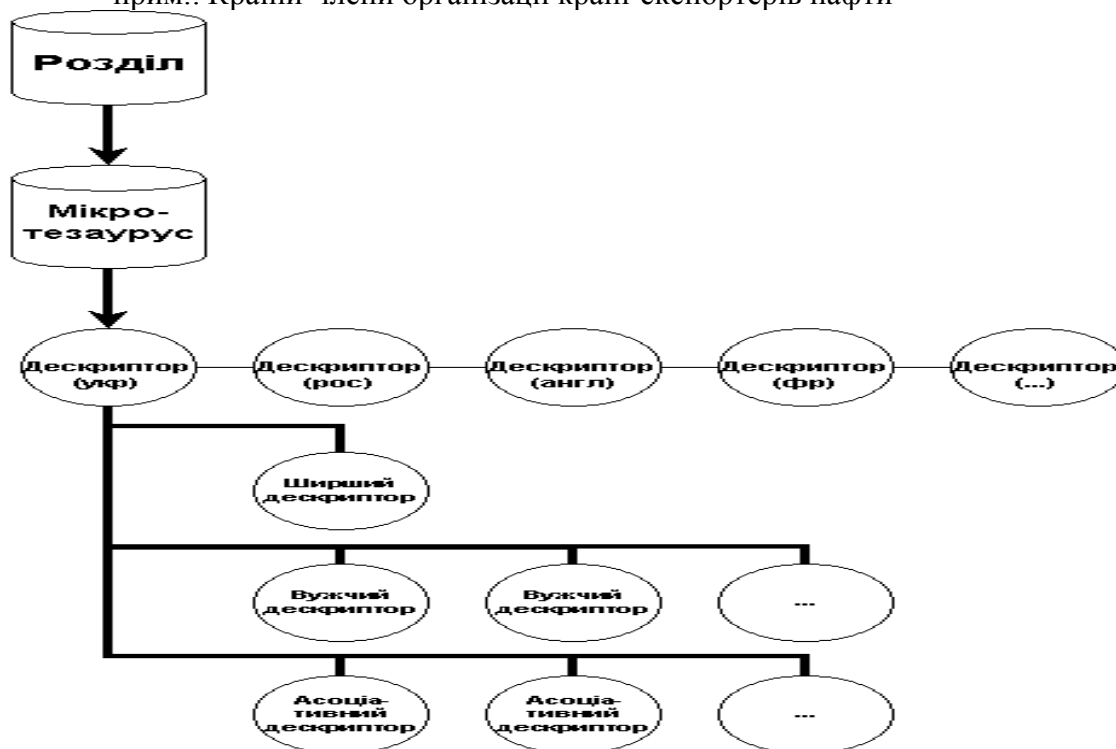


Рис. 4

– **приналежність до мікротезауруса** (зазначаються код мікротезауруса або коди декількох мікротезаурусів в розділах “ГЕОГРАФІЯ” та “МІЖНАРОДНІ ОРГАНІЗАЦІЇ”, до яких належить дескриптор)

Приклади: **ЗАХИСТ ПРАВ АКЦІОНЕРІВ**

мт: 2426 ФІНАНСУВАННЯ ТА ІНВЕСТУВАННЯ

УКРАЇНА

мт: 7206 ЄВРОПА І КОЛИШНІЙ РАДЯНСЬКИЙ СОЮЗ

мт: 7236 ПОЛІТИЧНА ГЕОГРАФІЯ

– **ієрархічні відношення** (об’єднують поняття “рід – вид” та “частка – ціле”. Для кожної пари дескрипторів, що зв’язані ієрархічними відношеннями, у дескрипторній статті ши-

ршого дескриптора вказується вузчий, а у дескрипторній статті вузчого – ширший. У випадку, коли має місце полієрархія в середині одного мікротезауруса, ширші дескриптори впорядковуються за абеткою);

Приклад: **БЕЛЬГІЯ**

мт: 7206 ЄВРОПА І КОЛИШНІЙ РАДЯНСЬКИЙ СОЮЗ

ш. ЗАХІДНА ЄВРОПА

мт: 7231 ЕКОНОМІЧНА ГЕОГРАФІЯ

ш. КРАЇНИ БЕНЛЮКСУ

ш. КРАЇНИ ЄС

ЗАХІДНА ЄВРОПА

мт: 7206 ЄВРОПА І КОЛИШНІЙ РАДЯНСЬКИЙ СОЮЗ

в БЕЛЬГІЯ

КРАЇНИ БЕНЛЮКСУ

мт: 7231 ЕКОНОМІЧНА ГЕОГРАФІЯ

в БЕЛЬГІЯ

КРАЇНИ ЄС

мт: 7231 ЕКОНОМІЧНА ГЕОГРАФІЯ

в БЕЛЬГІЯ

- **асоціативні відношення** (асоціативні відношення поєднують відношення “бути причиною”; ієрархічні відношення, коли полієрархія могла би бути, але заборонена і тому заміняється на асоціацію; відношення “супутності”, “близькості в просторі або часі”, “бути речовиною, із якої складається...”, “бути властивістю, особливістю”, “бути предметом дії, процесу, галузі знань”; відношення “локалізації”, “схожості, подібності” – коли два умовних синоніми мають статус дескрипторів, “антонімії”. На встановлення асоціативних відношень існують обмеження: дескриптори не можуть одночасно бути пов’язані ієрархічним та асоціативним зв’язками; дескриптори, які мають спільний ширший дескриптор не можуть бути пов’язані асоціативним зв’язком між собою. Для кожної з пар дескрипторів, що зв’язані асоціативними відношеннями, в статті кожного з двох дескрипторів подається посилання на асоціативно зв’язаний дескриптор; у дескрипторній статті, якщо асоціативний дескриптор належать до іншого мікротезауруса, то поряд з таким дескриптором вказується код його мікротезауруса або коди його мікротезаурусів для розділів “ГЕОГРАФІЯ” та “МІЖНАРОДНІ ОРГАНІЗАЦІЇ”).

Приклад: **ЕНЕРГЕТИЧНА ПОЛІТИКА**

мт: 6606 ЕНЕРГЕТИЧНА ПОЛІТИКА

а ЕНЕРГЕТИЧНІ РЕСУРСИ (5211)

а ЕНЕРГЕТИЧНІ ТЕХНОЛОГІЇ

ЕНЕРГЕТИЧНІ РЕСУРСИ

мт: 5211 ПРИРОДНЕ СЕРЕДОВИЩЕ

а ЕНЕРГЕТИЧНА ПОЛІТИКА

ЕНЕРГЕТИЧНІ ТЕХНОЛОГІЇ

мт: 6606 ЕНЕРГЕТИЧНА ПОЛІТИКА

а ЕНЕРГЕТИЧНА ПОЛІТИКА

КРАЇНИ ВИШЕГРАДСЬКОЇ ГРУПИ

мт: 7236 ПОЛІТИЧНА ГЕОГРАФІЯ

а УГОРЩИНА (7206+7231+7236)

У друкованій версії тезауруса EUROVOC дескриптори з позначкою семантичних відношень, які їх зв’язують, подаються в абетково-пермутаційному представленні.

Форми представлення української друкованої версії ППТ EUROVOC

Видання української версії EUROVOC здійснено у трьох томах: багатомовне представлення; тематичне представлення; абетково-пермутаційне представлення.

Багатомовне представлення. Багатомовне представлення містить:

- перелік розділів і мікротезаурисів EUROVOC українською, російською, англійською та французькою мовами. Розділи та мікротезауруси впорядковані за кодами EUROVOC та розташовані в чотири колонки;
- перелік дескрипторів EUROVOC українською, російською, англійською та французькою мовами з позначеннями кодів мікротезаурисів. Дескриптори розташовані за абеткою української мови в чотири колонки.

Тематичне представлення. У тематичному представленні дескриптори згруповані за розділами та мікротезаурусами з позначенням ієрархічних відношень. Заголовок мікротезауруса містить його код та назву. Кожен з мікротезаурисів починається з нової сторінки. Розділи та мікротезауруси подаються в порядку зростання їх кодів.

У середині мікротезауруса дескриптори розташовані у дві колонки. Дескриптори верхнього рівня (ті, що не мають ширших дескрипторів) розташовані за абеткою. Ієрархічні відношення між дескрипторами позначаються літерами v1, v2, v3..., де цифри вказують рівень ієрархії, в якій дескриптор з більшою цифрою є вужчим по відношенню до дескриптора з меншою, або до дескриптора верхнього рівня. Дескриптори, що знаходяться на одному рівні ієрархії впорядковані за абеткою.

Приклад: **ПРОДУКЦІЯ ГІРНИЧОДОБУВНОЇ ПРОМИСЛОВОСТІ**

- v1 БІТУМІНОЗНІ МАТЕРІАЛИ
- v1 КОШТОВНІ КАМЕНІ
- v1 МЕТАЛЕВА РУДА
 - v2 ЗАЛІЗНА РУДА
 - v2 РУДИ КОЛЬОРОВИХ МЕТАЛІВ
 - v3 БОКСИТ
- v1 НЕМЕТАЛЕВА РУДА
 - v2 АЗБЕСТ
 - v2 ПОТАШ

Дескриптор “**ПРОДУКЦІЯ ГІРНИЧОДОБУВНОЇ ПРОМИСЛОВОСТІ**” є дескриптором верхнього рівня і не має ширших дескрипторів. По відношенню до нього дескриптори “**БІТУМІНОЗНІ МАТЕРІАЛИ**”, “**КОШТОВНІ КАМЕНІ**”, “**МЕТАЛЕВА РУДА**”, “**НЕМЕТАЛЕВА РУДА**” та “**НЕРУДНІ МАТЕРІАЛИ**” є вужчими (перший рівень ієрархії). В той же час дескриптори “**ЗАЛІЗНА РУДА**” та “**РУДИ КОЛЬОРОВИХ МЕТАЛІВ**” є вужчими по відношенню до дескриптора “**МЕТАЛЕВА РУДА**” (другий рівень ієрархії). Аналогічно дескриптор “**НЕМЕТАЛЕВА РУДА**” теж має вужчі дескриптори. Нарешті дескриптор “**БОКСИТ**” є вужчим по відношенню до дескриптора “**РУДИ КОЛЬОРОВИХ МЕТАЛІВ**” (третій рівень ієрархії).

Абетково-пермутаційне представлення. При такому представленні за абеткою йдуть дескриптори, впорядковані як за своєю звичайною формою, так і інверсійною, за якою один і той самий дескриптор подається стільки разів, скільки в ньому значущих слів; при цьому кожне значуще слово виводиться наперед (спочатку перше слово переноситься в кінець словосполучення, потім – друге і так далі). На Рис. 5 зображені інверсійні форми для дескриптора, який складається з чотирьох значущих слів. Незначущі слова (наприклад, прийменники) відкидаються.

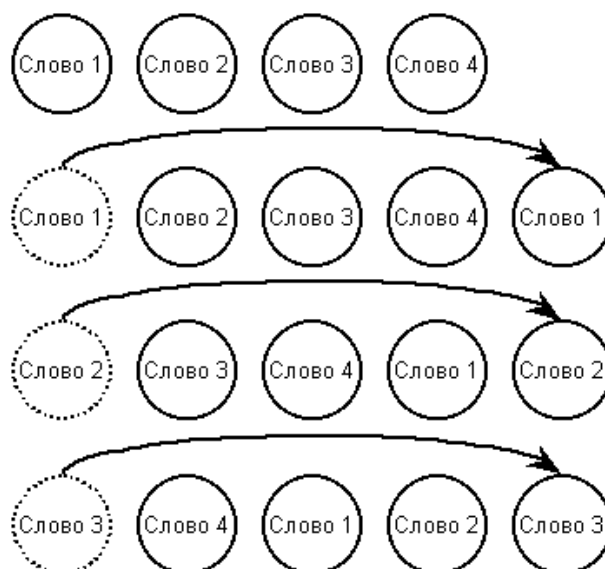


Рис. 5

Наприклад, дескриптор “СОЮЗ МІЖНАРОДНОГО СУДНОПЛАВСТВА ПО РЕЙНУ”. Відповідно, його інверсійні форми будуть:

- МІЖНАРОДНОГО СУДНОПЛАВСТВА ПО РЕЙНУ; СОЮЗ
- СУДНОПЛАВСТВА ПО РЕЙНУ; СОЮЗ МІЖНАРОДНОГО
- РЕЙНУ; СОЮЗ МІЖНАРОДНОГО СУДНОПЛАВСТВА

Кожна з цих форм буде знаходитись у відповідному місці алфавітного переліку дескрипторів.

Дескриптори у прямій формі подаються разом з тезаурусними статтями. Семантичні відношення всередині дескрипторної статті вказуються в такому порядку:

- прим.:** – примітка;
мт: – код і назва мікротезауруса;
ш – ширший дескриптор;
в – вужчий дескриптор;
а – асоціативний дескриптор.

У разі, якщо асоціативний дескриптор береться з іншого мікротезауруса, то після його назви в дужках подається код мікротезауруса-джерела.

Коди і назви мікротезаурусів у середині дескрипторних статей виділяються курсивом.

Приклад: **ПОЛЩЕЙСЬКИЙ КОНТРОЛЬ**

мт: 0431 ПОЛІТИКА І СУСПІЛЬНА БЕЗПЕКА
 ш СУСПІЛЬНИЙ ПОРЯДОК
 н ПРИКОРДОННИЙ КОНТРОЛЬ
 а ЗОВНІШНІЙ КОРДОН ЄС (1231)
 а ПОСВІДЧЕННЯ ОСОБИ (1231)

Дескриптори в інверсійній формі мають посилання на код свого мікротезауруса і подаються без тезаурусних статей.

Приклад: **КОНТРОЛЬ; ПОЛЩЕЙСЬКИЙ (0431)**

Якщо дескриптор належить до кількох мікротезаурусів, то ширші, вужчі та асоціативні дескриптори подаються після коду і назви мікротезауруса.

Приклад: **БІЛОРУСІЯ**

мт: 7206 ЄВРОПА І КОЛИШНІЙ РАДЯНСЬКИЙ СОЮЗ
 ш КОЛИШНІЙ СРСР
мт: 7236 ПОЛІТИЧНА ГЕОГРАФІЯ
 ш КРАЇНИ СНД

Електронна версія ПТТ EUROVOC

При адаптації ПТТ EUROVOC в Україні крім друкованого видання створюється електронна версія тезауруса. Електронна версія покликана не тільки спростити використання ПТТ EUROVOC при пошуку та індексації, й слугувати модулем для інформаційно-аналітичних правових систем в Україні.

Створення та впровадження електронної версії ПТТ EUROVOC спрямовано на інформаційну підтримку законотворчої діяльності Апарату Верховної Ради України, Комітетів Верховної Ради України, Інституту законодавства Верховної Ради України та надання можливості використання громадянами України свого права на доступ до чинного законодавства України та країн Європейського Союзу. Створення інтегрованої системи української версії ПТТ EUROVOC надасть змогу ознайомлення з українським законодавством міжнародної спільноти Європейського Союзу.

Електронна версія ПТТ EUROVOC призначена для:

- пошуку нормативно-правових актів України або міжнародних документів, які були ратифіковані Україною;
- аналізу чинного законодавства України та країн-членів Євросоюзу на предмет дублювання, суперечностей, прогалин та інших вад з метою їх подальшого усунення в законодавстві України та його адаптації до законодавства країн Європейського Союзу;
- підготовка до нових версій видань української версії тезауруса EUROVOC.

Метою створення електронної версії ПТТ EUROVOC є розробка єдиного лінгвістичного засобу індексування та пошуку нормативних документів, що утворюються в процесі законотворчої діяльності Верховної Ради України, а також забезпечення можливості використання тезауруса для міжнародного інформаційного обміну у сфері законодавства з використанням найсучасніших комп'ютерних та інформаційних технологій.

З цією метою електронна версія ПТТ EUROVOC повинна забезпечувати:

- збирання, опрацювання і систематизацію нормативно-правових актів України та пошук для аналізу міжнародних документів, які були ратифіковані Україною;
- створення та підтримку системи логічних зв'язків між документами;
- створення бази даних української версії тезауруса EUROVOC.

