

АНАЛІЗ ТА ОБҐРУНТУВАННЯ ВИБОРУ МЕТОДІВ ДОСЛІДЖЕННЯ ІНФОРМАЦІЙНОГО ПРОДУКТУ

© Вовк О.Б., 2014

Розглянуто методи дослідження інформаційного продукту, методи аналізу його показників, методи прогнозування майбутніх станів інформаційного продукту, окреслено основні складності, що виникають при цьому.

Ключові слова: інформаційний продукт, інформаційні технології, веб-сайт.

The article deals with methods of research of an information product, methods of analysis of its performances, methods of forecasting future states of an information product, outlines the major complexities involved in this.

Key words: information product, information technology, web-site.

Вступ. Постановка проблеми у загальному вигляді

Проектування якісних інформаційних продуктів (ІП) на основі технічних критеріїв не викликає труднощів, і для кожного класу ІП визначено типові проектні рішення.

Проте прогнозування майбутнього його стану на основі аналізу його минулого і сучасного, систематична інформація про якісні й кількісні характеристики цього інформаційного продукту, його поведінки в перспективі, врахування додаткових критеріїв викликає значні ускладнення, пов'язані з їх конфліктністю, повною або частковою схожістю критеріїв з різних груп тощо.

Тому важливим є вибір методів дослідження інформаційних продуктів, аналіз можливостей застосування їх до інформаційних продуктів загалом і веб-сайтів зокрема, та окреслення основних складностей, які виникають при цьому, задля прогнозування майбутніх станів інформаційного продукту.

Аналіз останніх досліджень та публікацій

Визначень інформаційного продукту є багато [1–5]. Зокрема Д.Г. Ейтутіс під “інформаційним продуктом” розуміє кінцевий результат процесу інформаційно-обчислювальної діяльності, наданий користувачу в матеріальній або нематеріальній формі [4].

Результатом такої діяльності є як фізичні об'єкти, так і послуги зі збору, оброблення інформації, забезпечення нею споживачів. Такий підхід дає можливість встановити ціну на інформаційний продукт, яка складається з вартості витрачених ресурсів і прибутку.

Інший дослідник, М.О. Ясева, розглядає інформаційний продукт у контексті теорій трудової вартості та граничної корисності [2]. Зокрема, автор зазначає, що сьогодні не існує єдиної теорії інформаційного продукту. Необхідність існування такої теорії викликана активізацією ринку інформаційних товарів та послуг, а також ризиками інформаційного простору.

З позиції виробника інформаційний продукт – це сукупність даних, сформована для поширення в матеріальній або нематеріальній формі.

В статті поняття інформаційного продукту визначено так:

інформаційний продукт (ІП) (англ. information product) – документовані дані, підготовлені відповідно до потреб користувачів і подані у вигляді товару.

З погляду інформаційних технологій інформаційними продуктами є програмні продукти, бази і банки даних, веб-сайти тощо.

Загалом ІП складається з інформаційних ресурсів (IR), інтелектуальних ресурсів (HR), метаданих про цей ІП (M): $I_p = \langle IR, HR, M \rangle$.

Інформаційний ресурс (ІР) (IR) – сукупність структурованих даних для отримання достовірної інформації, фіксована на матеріальному носії [5].

Інтелектуальні ресурси (інформаційно-інтелектуальні ресурси) (ІнР) (*HR*) – сукупність об'єктивних форм інформації чи знань, отриманих в результаті інформаційно-технологічних, науково-виробничих, організаційно-управлінських, маркетингових, фінансових, юридичних тощо методів, технологій, засобів, різних форм подання завдяки інтелектуальній праці [7].

Інтелектуальними ресурсами є програмні модулі, структурні схеми, технічні характеристики ІІІ тощо.

З погляду інформаційного суспільства поділ ІІІ відбувається за:

- вирішенням актуальних проблем суспільства;
- інформаційною підтримкою суспільства;
- навчанням;
- дозволяям.

Інформаційні продукти і послуги становлять основу ринку ІІІ та охоплюють:

- електронні матеріали;
- друковані матеріали;
- комбінації друкованих та електронних матеріалів (зокрема – на DVD чи CD);
- аудіофайли (а також аудіокниги);
- відеофайли;
- комп'ютерні програми, БД, БЗ;
- високоякісні цифрові фотографії;
- веб-сайти;
- музичні твори;
- мистецькі твори;
- консалтингові продукти, де учасники можуть отримати доступ до потоку інформації [9].

Апробацію розглянутих в статті методів дослідження інформаційних продуктів здійснено на такому виді інформаційних продуктів, як веб-сайт.

Веб-сайт – це інформаційний продукт, що є результатом функціонування інформаційної технології (комбінація програмних засобів і даних) [11].

Веб-сайт має власну (“фізичну”) адресу, довідкові та регулярні дані, файли тощо.

Нижче наведено специфікацію вимог до веб-сайтів, структуровану відповідно до моделі якості:

- точність надання інформації;
- зрозумілість структури сайту;
- продуктивність сайту;
- швидкість надання інформації;
- безпека зберігання та контролю над даними;
- безпека роботи протягом визначеного періоду часу;
- надійність сайту;
- доступність;
- ефективність;
- наявність системи навігації;
- зрозумілість структури сайту;
- естетичне оформлення сайту;
- наявність системи для ведення статистики отриманих послуг;
- відповідність сайту галузевим чи міжнародним стандартам;
- економія часових ресурсів;
- економія фінансових затрат;
- наявність засобів наповнення інформаційного контенту [11].

У наведеній специфікації вимог до веб-сайтів, які сформовані на основі моделі якості, в уніфікованій формі подано потреби замовника та користувачів.

Прогнозування – процес передбачення майбутнього стану предмета чи явища на основі аналізу його минулого і сучасного, систематична інформація про якісні й кількісні характеристики поведінки цього предмета чи явища в перспективі [10].

У міжнародній практиці поняття “передбачення” переважно застосовують до технологічного сегмента, що привело до появи терміна “технологічне передбачення” [14].

Узагальнено прогнозування поділяється на:

1. Економічне:

1.1. Ланцюг постачання – прогнозування застосовується в ланцюгу постачання для забезпечення клієнтів компаній потрібним продуктом у точний час.

1.2. Бізнес планування – частина підготовки та розроблення бізнес-планів.

2. Природниче:

2.1. Прогноз погоди.

2.2. Метеорологія.

2.3. Передбачення землетрусів.

3. Технічне та технологічне.

3.1. Технічне прогнозування.

3.2. Технологічне прогнозування.

3.3. Логістичне прогнозування [15].

У статті розглянуто технічне прогнозування.

Процес прогнозування ґрунтується на певних принципах, основними з них є:

- цілеспрямованість – змістовий опис поставлених дослідницьких завдань;
- системність – побудова прогнозу на підставі системи методів і моделей, що характеризуються певною ієрархією та послідовністю;

- наукова обґрунтованість – усебічне врахування вимог об’єктивних законів розвитку, використання світового досвіду;

- багаторівневий опис – опис об’єкта як цілісного явища і водночас як елемента складнішої системи;

- інформаційна єдність – використання інформації на однаковому рівні узагальнення й цілісності ознак;

- адекватність об’єктивним закономірностям розвитку – виявлення та оцінка стійких взаємозв’язків і тенденцій розвитку об’єкта;

- послідовне усунення невизначеностей – процедура просування об’єкта від виявлення цілей та умов, що склалися, до визначення можливих напрямів розвитку;

- альтернативність – виявлення можливості поведінки об’єкта за умови різних траєкторій, різноманітних взаємозв’язків і структурних співвідношень [16].

Формування цілей

Метою дослідження є обґрунтування вибору методів дослідження інформаційного продукту та аналіз методів прогнозування його майбутнього стану на основі розгляду минулого та сучасного його функціонування з використанням систематичної інформації про якісні й кількісні характеристики цього інформаційного продукту.

Аналіз отриманих наукових результатів.

Методи дослідження інформаційних продуктів

Методи теорії інформації. Теорія інформації вивчає кількісні закономірності, пов’язані з отриманням, опрацюванням, передачею, зберіганням інформації.

На інформаційний продукт як об’єкт, що містить інформацію, також поширюються ці закономірності, оскільки оцінка кількості інформації ґрунтується на законах теорії вірогідності й визначається через вірогідність подій. Це означає, що інформація має цінність тільки тоді, коли користувач довідується з неї про подію випадкового характеру, тобто результат події завчасно не відомий.

Кількісну міру інформації можна визначити як: $M(I) = \log_2(A)$, де $M(I)$ – кількісна міра інформації; I – інформація; A – адитивність (властивість, яка полягає в тому, що значення величини, яка відповідає цілому об'єкту, дорівнює сумі значень величин, що відповідають його частинам, незалежно від того, як поділено об'єкт).

Для інформаційних продуктів (наприклад, для веб-сайта) адитивність притаманна кожному окремому елементу, наприклад: БД, БЗ тощо.

Що більше інформації в ІІ, що цікавить користувача, має випадковий характер, то більше інформації містить цей інформаційний продукт і то цінніший він.

Наприклад, існує деякий інформаційний продукт (веб-сайт) X , який випадково може опинитися в тому чи іншому стані (наприклад, корисний або некорисний, корисний частково, може бути корисним через якийсь час). Отже, цьому ІІ притаманний певний ступінь невизначеності, яка характеризується не тільки кількістю станів об'єкта, але і вірогідностями станів.

Загалом, деякий ІІ X може набувати якусь скінченну кількість станів:

$$x_1, x_2, x_3, \dots, x_n$$

з вірогідностями $P_1, P_2, P_3, \dots, P_n$:

$$p_i = P(X \approx x_i),$$

вірогідність того, що інформаційний продукт X набуде стану x_i , тобто:

$$X \approx x_i: \sum_i^n p_i = 1.$$

Як міра апіорної невизначеності в теорії інформації використовується ентропія. Для ІІ ентропія розглядається як міра невизначеності його стану в момент часу t_i . Під ентропією ІІ розуміють суму добутків вірогідностей різних станів ІІ на логарифми цих вірогідностей, взяті з протилежним знаком:

$$H(X) = -\sum_i^n p_i \log_2 p_i.$$

Нехай є деякий сайт X , над яким здійснюється дослідження. Оцінимо інформацію, що буде отримана в результаті того, що стан ІІ X стане повністю відомим.

До отримання даних (відомостей) апіорі ентропія об'єкта була $H(X)$. Після отримання даних (відомостей) стан об'єкта повністю визначився, тобто ентропія тепер дорівнює 0.

Якщо I_x – інформація, отримана в результаті визначення стану об'єкта X , то вона дорівнює зменшенню ентропії:

$$I_x = H(X) - 0 \text{ або } I_x = H(X).$$

Тобто кількість інформації, отриманої під час з'ясування стану будь-якого ІІ, дорівнює ентропії цього об'єкта.

Використовуючи формулу для ентропії, отримаємо таке представлення кількості інформації:

$$I_x = -\sum_i^n p_i \log_2 p_i.$$

У процесі отримання даних (відомостей) про ІІ невизначеність може бути зменшена. Якщо обсяг визначених даних, наділених ознаками повноти, несуперечливості, актуальності, збільшиться, тоді ІІ буде змістовнішим, оскільки більшою буде інформація про досліджуваний ІІ, і як наслідок, – меншою невизначеність. У цьому випадку кількість інформації характеризуватиметься зменшенням ентропії ІІ, для уточнення стану якого призначені дані (відомості).

Такий підхід дає змогу застосовувати в процесі оцінки майбутнього потенційного функціонування інформаційного продукту ентропійні характеристики, в які входить поняття “активної оцінки” ІІ – оцінки, яка дозволяє кількісно (у відсотках) оцінити можливі варіанти поведінки ІІ.

Для активної оцінки інформаційного продукту можна використовувати шкали з різними коефіцієнтами градації оцінок N , наприклад:

$N=2$: прийнятний, неприйнятний;

N=3: прийнятний, задовільний, неприйнятний;

N=4: високоефективний, прийнятний, задовільний, неприйнятний.

Інакше кажучи, активна оцінка не тільки характеризує ІІ, але й дає змогу здійснювати прогноз щодо нього. Одним з варіантів такої оцінки може бути підхід інвертування простору групової оцінки (рис. 1) в ентропійний простір (рис. 2).

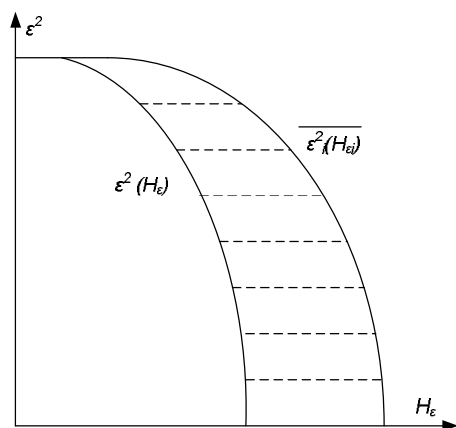


Рис. 1. Середній кількісний простір групової оцінки ІІ

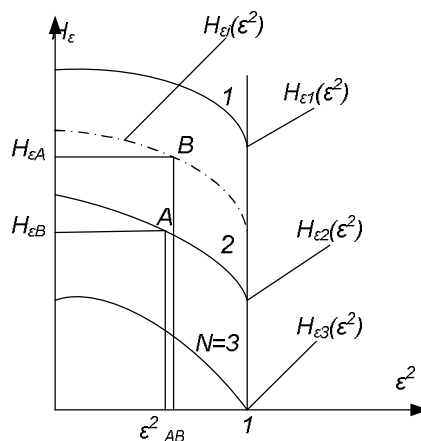


Рис. 2. Ентропійний простір оцінки ІІ

Проте інколи за надлишковості даних з'являється інформаційний шум, який призводить до неповноти або спотворення отриманої інформації, погіршуючи її якість. У цьому випадку збільшення даних не веде до зменшення невизначеності.

Методи аналізу показників інформаційних продуктів

Метод синтаксичного та семантичного аналізу. Синтаксичний аналіз встановлює найважливіші параметри інформаційних потоків, зокрема кількісні характеристики, для вибору комплексу технічних програмних засобів збору, накопичення, реєстрації, обробки, передачі та зберігання інформації.

Семантичний аналіз досліджує інформацію з погляду змісту її окремих елементів та дає можливість проводити змістовий аналіз інформації на основі вивчення значень одиниць мови з однозначним розпізнаванням повідомлень, аналіз зміни значень одиниць мови у їх різних сполученнях. Проте це проблема достатньо складна і є основою таких додатків, як автоматичні відповіді на питання з бази даних або відновлення інформації шляхом необмежених запитів природної мови.

Одним з варіантів цього методу є метод “гіпертекстової розмітки”, оснований на призначенні кожній одиниці інформації однієї (чи більше) функції чи зв'язків (залежно від виконуваної функції). В ньому фрагменти інформаційного продукту розглядаються як зв'язані (наприклад, поняттями, ієрархіями, досвідом, поведженням тощо).

Зв'язки між такими фрагментами в інформаційному продукті будуються через “гіперпосилання” (англ. – hyperlinks) [8]. Об'єкти (текст, графіка, зображення, аудіо та відео), зв'язані так, можуть бути поєднані в гіпермедійні бази даних. Тому цей метод потенційно збільшує міру подібності. Проте недоліком цього методу є те, що не кожен ІІ можна розкласти на фрагменти або не всі елементи між собою можна пов'язати “гіперпосиланнями”.

Метод евристичних функцій. Основне призначення методу евристичних функцій – побудова моделей процесів прийняття рішень будь-якої нової задачі, які передбачають такі варіанти:

- сліпий пошук, який ґрунтується на методі “спроб та помилок”;
- варіант “лабіринту”, в якому розв'язання задачі розглядається як лабіринт, а процес пошуку рішення – як його проходження;
- структурно-семантичний варіант, який є найзмістовнішим і відображає семантичні відношення між об'єктами, що входять в задачу.

Недоліком евристичного методу є те, що процес розв'язання задачі відбувається тільки в конкретній предметній області.

Для дослідження інформаційних продуктів та прогнозування їх поведінки цей метод використовується під час розроблення сценаріїв майбутніх станів ІІ.

Метод соціального моделювання. Метод соціального моделювання використовується для дослідження процесів ІІ з позиції соціуму [10]. Недоліком цього методу є певна абстрактність.

Використання методу соціального моделювання для прогнозування майбутніх станів веб-сайта полягає в моделюванні його позиції в механізмах збору, обробки та використання інформації відносно конкретної соціальної групи, управлінні ним тощо.

Кібернетичний метод (інша назва – метод єдності прямих та зворотних інформаційних зв'язків) є доволі цінним джерелом інформації під час дослідження будь-якого об'єкта.

Зокрема, прямий інформаційний зв'язок визначає поведінку об'єкта залежно від певних впливів, а зворотний зв'язок визначає вибір впливу залежно від попередньої поведінки цього об'єкта. Недоліком цього методу є обов'язкова наявність постійного зв'язку між прямим та зворотним механізмами.

Наприклад, якщо в механізмі використання інформаційного продукту є тільки прямий зв'язок, то постає задача його ефективного використання без наявної інформації, оскільки зворотна інформація про стан інформаційного суспільства і реальний стан управління таким ІІ не надається. В такій ситуації відсутня друга складова контуру регулювання, яка називається зворотним інформаційним зв'язком. За таких умов немає адекватного регулювання ІІ.

В ситуації одномоментного зв'язку керування ним взагалі неможливе. Суб'єктами впливу в такій системі є користувачі ІІ – з одного боку та власники (чи розробники) – з іншого (рис. 3).

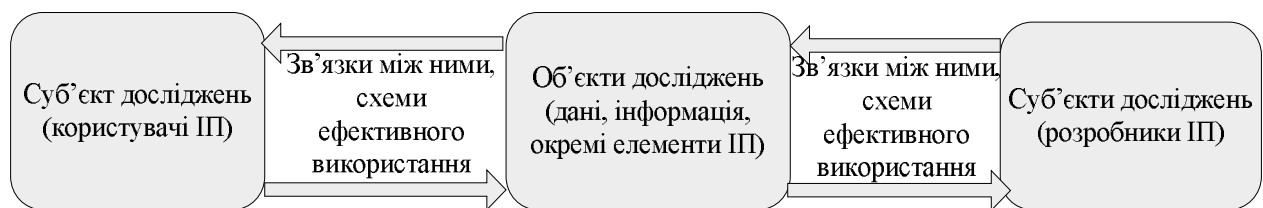


Рис. 3. Схема дослідження ІІ кібернетичним методом

Під час дослідження ІІ як кібернетичної системи виділяють переважно такі задачі:

- визначення сфери використання (функція корисності);
- координація та управління ІІ (з урахуванням масиву інформації, отриманої з різних джерел);
- визначення впливу можливих факторів ризику на життєздатність ІІ;
- досягнення максимальної ефективності використання ІІ.

Системний метод дає змогу виявити властивості, які виникають в результаті об'єднання окремих елементів в одну групу, одним зі способів узагальнення емпіричних фактів.

За системного підходу до аналізу ІІ важливими є:

- сфера використання;
- набір інтелектуальних та інформаційних ресурсів;
- зв'язки з іншими ІІ (гіперпосилання);
- послідовності зміни майбутніх станів ІІ.

Методи прогнозування майбутнього стану ІІ

Методи прогнозування формуються на основі фіксації структурованих за певними ознаками даних про об'єкт прогнозування.

Залежно від джерел інформації, технології її обробки та одержаних результатів методи прогнозування поділяються на групи [13]:

- I) фактографічні;
- II) евристичні.

I. Фактографічні методи прогнозування ґрунтуються на використанні фактичних матеріалів, що детально характеризують зміни в часі всієї сукупності чи окремих ознак (показників) об'єкта прогнозування. Основними в цій групі є методи:

- 1) кореляційно-регресійних моделей;
- 2) екстраполяції;
- 3) функцій.

Прогнозування з використанням кореляційних моделей (методів) полягає у пошуку математичних формул, що характеризують статистичний зв'язок одного показника з іншим (парна кореляція) або з групою інших (множинна кореляція). Зокрема, методи кореляційних і регресійних моделей широко застосовуються в економічному прогнозуванні.

Обов'язковою умовою можливості та доцільності застосування таких методів є встановлення ступеня надійності кореляційних формул на підставі логічного аналізу достатності статистичної вибірки (масиву даних).

Для інформаційного продукту форма взаємозв'язку прогнозованих явищ з іншими явищами та процесами описується регресійним рівнянням типу $y = f(x_1, x_2, \dots, x_n)$. А значення прогнозованого показника визначається способом підстановки в таке рівняння значення ознак (критеріїв) та оцінки очікуваного середнього значення результативної ознаки.

У процесі розв'язання кореляційних і регресійних рівнянь здійснюється пошук кількісних значень параметрів вихідних залежностей, які можна визначити за допомогою **методу найменших квадратів**, суть якого полягає у мінімізуванні суми квадратичних відхилень між величинами, що спостерігаються, та відповідними величинами, розрахованими за підібраними рівняннями зв'язку.

Суть **методу екстраполяції для задач прогнозування майбутнього стану інформаційного продукту** полягає в поширенні закономірностей розвитку об'єкта в минулому на його майбутнє, в якому за основу взято припущення про незмінність чинників, що визначають розвиток об'єкта дослідження.

Метод функцій ґрунтується на використанні так званих автокореляційних функцій, що виражається взаємним зв'язком між сусідніми членами часового ряду. Стосовно інформаційних продуктів процес прогнозування з використанням автокореляційних функцій полягає у виконанні двох послідовних дій:

- перша – формулювання завдання прогнозування й визначення критеріїв його вирішення;
- друга – використання часового ряду, який відображає процес розвитку параметрів об'єкта в часі, визначається прогнозована величина на перспективний період за умови мінімізації середньоквадратичних похибок передбачення.

Евристичні методи прогнозування передбачають використання методів прогнозування за допомогою логічних прийомів і методичних правил теоретичних досліджень.

Евристичні методи поділяються на інтуїтивні та аналітичні. З-поміж основних методів першої підгрупи виокремлюють методи експертної оцінки, а другої – методи морфологічного аналізу, побудови “дерева цілей”, інформаційного моделювання, оптимізації.

В ситуації, коли статистичної інформації недостатньо або вона непридатна, для прогнозування певних явищ необхідно використовувати **метод експертних оцінок**. В його основу покладено спосіб збирання необхідної інформації шляхом опитування експертів. Експертне опитування треба провести так, щоб можна було отримати:

- кількісно однозначні відповіді на запитання, що пропонуються експерту;
- формалізовані відомості щодо характеру критеріїв, ступеня впливу кожного з критеріїв;
- кількісно визначену експертом оцінку рівня його знання про об'єкт дослідження, що пропонується для аналізу та висновків.

Згідно з цим методом застосовують підхід індивідуальної або групової оцінки. У разі індивідуальної оцінки кожний експерт дає незалежну оцінку в аналітичному вигляді. Групові оцінки ґрунтуються на колективній роботі експертів та одержанні сумарної оцінки від усієї групи експертів, яких залучено.

Метод морфологічного прогнозування ґрунтується на використанні комбінаторики, тобто дослідженні всіх можливих варіантів, згідно із закономірностями побудови (морфології) об'єкта прогнозування, що вивчається та аналізується.

Метод побудови “дерева цілей” застосовується в прогнозуванні з метою поділу основних завдань на підзавдання і створення системи “виважених” за експертними оцінками зв'язків.

Специфічним методом прогнозування є **метод інформаційного моделювання**. Він ґрунтується на характерних особливостях масових потоків інформації, створюючи умови для прогнозування розвитку конкретних об'єктів на підставі таких джерел інформації, які містять необхідні, логічно впорядковані документи в певній послідовності.

Існують й інші методи прогнозування, зокрема, випереджувальні методи, до яких, наприклад, належать методи патентної експертизи. Для відбору чинників прогнозованої моделі та побудови системи зв'язків широко використовуються матриці взаємовпливу і теорія графів.

До підгрупи статистичних методів прогнозування належать методи: інтерполяції, кривих зростання та огинаючих кривих, формування сценаріїв, а до підгрупи аналітичних – методи аналогій тощо.

Узагальнюючи все вищесказане, можна стверджувати, що під час моделювання та прогнозування майбутніх станів інформаційного продукту слід використовувати комбінацію розглянутих методів. Це дасть змогу протиставити нечітким об'єктам, що мають неконкретні значення, чітко визначені змодельовані елементи, з виділенням простих представлень і фіксацією його елементів.

Зокрема, для прогнозування майбутніх станів ІІ на практиці доцільно використовувати такі підходи:

- системно-елементарний – опис елементів інформаційного продукту;
- системно-структурний – визначення структури ІІ і взаємодії компонентів, що її утворюють;
- системно-функціональний – визначення сфери застосування інформаційного продукту;
- системно-комунікаційний – побудова взаємозв'язку окремого ІІ з іншими в певній сфері застосування;
- системно-інтеграційний – представлення факторів впливу на ІІ та механізмів його вдосконалення;
- системно-історичний – встановлення зв'язку між різними етапами життєвого циклу ІІ в різних часових проміжках.

Крім того, для прогнозування майбутніх станів ІІ, окрім зазначених підходів, слід використати математичні методи, які реалізувати за два етапи:

- перший – дослідження за допомогою математичних методів;
- другий – математична обробка результатів досліджень.

Перший етап, використовуючи штучно створену мову, формує розгорнуті характеристики ІІ і закони їх використання. За всієї нескінченної множини характеристик кожна з них має єдине знакове представлення, а отже, кожному етапів присвоюється власна назва, що дає змогу певні дії записати особливою мовою.

На другому етапі – математичної обробки результатів досліджень – пропонується використовувати метод математичної логіки. Це дозволить описати всі процеси універсальною мовою та отримати:

- 1) розклад всіх понять на найпростіші; кількість таких понять повинна бути невеликою (подібно до того, як, в математиці складені числа розкладаються на добуток простих множників);
- 2) усі можливі комбінації простих понять, які в загальному утворюють множину складних (при цьому, хоч кількість простих понять мала, згідно з правилом комбінаторики, множина складних може бути нескінченною);
- 3) алфавіт, у якому кожному з понять відповідає особливий (конкретний) символ;
- 4) уніфіковані відношення між поняттями та правила їх використання.

Таке моделювання дає змогу виявити логічні (зокрема латентні) властивості або процеси, що перебувають у прихованому (невидимому) стані і не проявляють себе ніяк. Стосовно

інформаційних продуктів латентність може виражатись затримкою чи очікуванням відповіді, яка збільшує реальний час відгуку на ІІ порівняно з очікуваним та може бути зменшена прогнозуванням майбутніх станів, які в цей момент є неочевидними.

Хоча недоліком такого підходу є гіпотетичність та ідеалізація підрахунку, проте наведена формалізація дозволяє використати його для узагальнювального моделювання майбутнього стану ІІ.

Зокрема, логічну структуру інформаційного продукту можна подати у такому вигляді:

$$Surm_{IP} \rightarrow Dis \rightarrow Rez \rightarrow Prog,$$

де $Surm_{IP}$ – припущення щодо майбутнього стану ІІ; Dis – найкраще поточне рішення стосовно використання ІІ; Rez – результат; $Prog$ – прогнозування майбутніх станів ІІ (з урахуванням характеристик та факторів ризику).

Результатом прогнозування є ймовірні сценарії поведінки інформаційного продукту в майбутньому.

Надзвичайно важливим моментом для прийняття рішення є оцінка ймовірного прогнозу.

Серед множини технічних можливостей реалізації запропонованого підходу існує декілька, які є простими, але, водночас дають безпомилкову оцінку прогнозу об'єкта (в нашому випадку інформаційного продукту) щодо його подальшої поведінки, зокрема це:

- середнє абсолютне відхилення (англ. – Mean Average Deviation (MAD));
- сума помилок прогнозу зростаючим підсумком (англ. – Running Sum of Forecast Error (RSFE));
- сигнал відслідковування (англ. – Tracking Signal (TS)).

Висновки

Описано вибір математичних методів прогнозування майбутніх станів інформаційного продукту на основі розгляду минулого та сучасного його функціонування з використанням систематичної інформації про якісні й кількісні характеристики цього інформаційного продукту.

Проаналізовано технічні можливості реалізації кожного з методів, окреслено основні складності, які виникають при цьому.

Це дасть змогу в перспективі врахувати додаткові вимоги, що можуть ставитися до інформаційного продукту та викликати значні ускладнення, пов'язані з їх конфліктністю, повною або частковою схожістю з вже наявними тощо.

1. Юрьева Т. Ю. *Словарь информационных продуктов и услуг [Текст] / Т. Ю. Юрьева. – Кемерово : РОСТИКС, 2006. – С.50.* 2. Ясева М.О. *Суб'єкти механізму формування вартості інтелектуального продукту: економічні інтереси, цілі та мотивація // Економіка: проблеми теорії та практики: зб. наук. праць. – Вип. 259: В 3 т. – Т. I. – Дніпропетровськ: ДНУ, 2010. – С. 212-220.* 3. *Classification of information products. [Електронний ресурс]. – Режим доступу: http://en.wikipedia.org/wiki/Classification_of_information_products. – [Назва з екрана].* 4. *Different kinds of information products [Електронний ресурс]. – Режим доступу: http://www.pandecta.com/information_products.html. – [Назва з екрана].* 5. *Information is the one product that meets every one of those criteria. [Електронний ресурс] / John Thornhill. – Режим доступу: <http://www.skipmcgrath.com/products/info-product.shtml>. – [Назва з екрана].* 6. *Types of Information Products [Електронний ресурс]. – Режим доступу: <http://ezinearticles.com/?Types-of-Information-Products>. – [Назва з екрана].* 7. *Інтелектуальний ресурс [Електронний ресурс]. – Режим доступу: http://uk.wikipedia.org/wiki/Інтелектуальний_ресурс.* 8. *Weinreich Harald. The look of the link – concepts for the user interface of extended hyperlinks. p. 19. doi:10.1145/504216.504225. [Електронний ресурс] / Weinreich Harald, Hartmut Obendorf, Winfried Lamersdorf. – Режим доступу: <http://portal.acm.org/citation.cfm?id=504225>. – [Назва з екрана].* 9. *Shannon Claude. A Mathematical Theory of Communication [Електронний ресурс] / Claude Shannon, Warren Weaver // Bell System Technical Journal. – № 27 (July and October). – P. 379–423, 623–656. – Режим доступу: <http://plan9.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>. – [Назва з екрана].* 10. *Белл Д. Грядущее постиндустриальное общество. Опыт социального прогнозирования [Текст] / Белл Д.;*

пер. с англ. под ред. В.Л. Иноземцева. – М.: Academia, 1999. – 783 с. 11. *The network of everything // Research*eu result supplement.* – № 11 (January 2009). – Р. 30. 12. Ваниянц С. В. Информационный ресурс в экономической сфере [Текст] / С.В.Ваниянц. – М.: Прогресс, 2006. – С.56–59. 13. Василенко В.А. Теорія і практика розробки управлінських рішень [Текст] / Василенко В.А. – К.: ЦУЛ, 2003. – 176 с. 14. Кедровская Л. Г. Номенклатура информационных услуг [Текст] / Л.Г. Кедровская, А. И. Мишеллидзе, Ю. Н. Ухин. – СПб.: ИПКИР, 2005.– 180 с. 15. Каган М.С. Избранные труды в VII томах. Том I. Проблемы методологии. Системное рассмотрение основных способов группировки [Текст]. – Санкт-Петербург: ИД “Петрополис”, 2006. – 356 с. 16. Фрэнк Уэбстер. Теории информационного общества [Текст] / Фрэнк Уэбстер; пер. с англ. Е. Л. Вартановой. – М.: Аспект пресс, 2004. – 400 с.

УДК 004.89

Р.В. Вовнянка¹, Д.Г. Досин², В.В. Ковалевич²

¹Национальный университет “Львівська політехніка”,
кафедра інформаційних систем та мереж,
²Фізико-механічний інститут ім. Г.В. Карпенка

МЕТОД ВИДОБУВАННЯ ЗНАНЬ З ТЕКСТОВИХ ДОКУМЕНТІВ

© Вовнянка Р.В., Досин Д.Г., Ковалевич В.В., 2014

Запропоновано метод, алгоритм і засоби для виділення знань з природномовного тексту. Показано, що такий алгоритм має бути багатостадійним і містити ієрархічну кількарівневу процедуру розпізнавання понять, зв'язків, предикатів та правил, які в результаті вносяться до онтології.

Ключові слова: онтологія, навчання онтологій, інтелектуальний агент, база знань, текстовий документ.

In the paper a method an algorithm and tools for selection of knowledge from a text document are suggested. It is shown that this algorithm has to be multistage and involve hierarchical procedure of concepts recognition of relations, predicates and rules which are introduced into the resulting ontology.

Key words: ontology, learning ontologies, intelligent agent, knowledge base, text document.

Вступ. Постановка проблеми у загальному вигляді

Поняття “знання” належить до галузі наукових досліджень методів і засобів прийняття оптимальних рішень. В процесі набуття знань через навчання суб'єкт прийняття рішень використовує доступну йому інформацію для побудови оптимальної стратегії прийняття рішень. Інформація у нашому розумінні набуває статусу знань саме тією мірою, якою вона допомагає носію цієї інформації вирішити його завдання і може бути числово оцінена як виграш від її використання для прийняття рішень у процесі досягнення відповідних цілей. Належно організовану і впорядковану сукупність знань інтелектуального агента називають базою знань. Система впорядкування знань у такій базі знань формально-логічно сформульована у її онтології, визначеній як “експліцитна специфікація концептуалізації” [1], тобто явне формальне означення понять і допустимих семантичних зв'язків між ними.

Суть методу видобування знань з природномовного текстового документа, інакше кажучи, розпізнавання змісту текстового документа, полягає у побудові плану (стратегії) діяльності інтелектуального агента – інформаційної моделі суб'єкта розпізнавання або уточнення такого плану на підставі даних, виділених у текстовому документі, що розпізнається. Тут вважаємо план конкретною реалізацією оптимальної стратегії розв'язання деякої задачі, що стоїть перед інтелектуальним агентом у межах заданої проблемної області.