

УДК 027. 081 : 004

Віктор ХЛАНЬ,

провід. наук. співроб. ІДПДБ

Антон ВІТУШКО,

провід. інженер відділу технологій електронної обробки інформації НБУВ

Формування моделі користувача в інтелектуальній інформаційній системі

Досліджується тема інтелектуалізації інформаційного пошуку шляхом створення методів адаптування до інформаційних потреб аналітика завдяки моделі користувача, аналізується процес формування моделі користувача в інтелектуальній інформаційній системі.

У зв'язку з активізацією процесів інформатизації суспільства принципово змінюється роль бібліотек у сучасному світі. Бібліотека стає інтелектуальним посередником між людиною та інформацією. І потреба в таких послугах зростає, оскільки з величезною швидкістю збільшується обсяг доступної інформації, розмаїтість її подання й засобів роботи з нею. Все це істотно ускладнює для користувача завдання пошуку потрібної інформації і її подальшу обробку.

Завдяки більш широким інформаційним можливостям бібліотек виникає необхідність у наданні нових видів сервісів. Однією з важливих складових цих сервісів є інформаційно-аналітичне забезпечення.

Ефективне розв'язання аналітичних завдань потребує наукової роботи та подальшого впровадження в практичну діяльність відповідних підрозділів інтелектуальних технологій (ІТ), за допомогою яких можна опрацьовувати значні обсяги неструктурованої інформації.

Зокрема, суттєвого значення набувають ІТ, які забезпечують ефективну реалізацію усіх видів інтелектуальної діяльності та, у першу чергу таких, що складно формалізуються. На тлі впровадження ІТ, призначених для спрощення виконання аналітиком рутинних операцій та окремих найпростіших функцій, необхідно передбачити інтенсифікацію і підвищення ефективності інтелектуальних видів діяльності за рахунок комплексного використання ІТ, спрямованих на стимулювання евристичного, асоціативного й образного мислення аналітиків.

Вказане потребує накопичення та подальшого використання знань щодо процесу трансформації інформаційної потреби аналітика в запиті до інтелектуальної інформаційної системи (ІС). Ґрунтуючись на отриманих знаннях, система має сформувати гіпотезу щодо необхідної

аналітику інформації та здійснити генерацію запиту на її одержання.

Оскільки інформація, з якою працює аналітик, не формалізована, у роботі пропонується використовувати інтелектуальну інформаційну систему, основу якої становитиме семантична мережа. Водночас, з метою якісного задоволення інформаційної потреби аналітика, при опрацюванні документальних матеріалів щодо досліджуваної предметної галузь вважається за доцільне враховувати інформацію про перебіг виконання інформаційною системою запитів користувача. Останнє потребує створення та подальшого використання в роботі інтелектуальної інформаційної системи моделі користувача та моделі його поведінки під час пошуку необхідної інформації [1].

Під моделлю користувача будемо розуміти наявні в ІС знання про аналітика, які використовуються для відбору необхідної інформації та подальшої її видачі (візуалізації). Модель користувача описує (на поточний стан) знання та вміння конкретного аналітика й вважається «ідеальною» моделлю знань про аналітика, яка враховує знання про предметну галузь, типові помилки та когнітивні механізми.

Зокрема, модель користувача має враховувати: загальні характеристики аналітика, які не залежать від досліджуваної предметної галузі (наприклад: вік, стать, швидкість реакції, освіта, спеціальність тощо); ставлення аналітика до наданої інформації (документів), а саме: наскільки глибоко й повно він володіє знаннями і/або вміннями щодо досліджуваної предметної галузі; етапи взаємодії між системою та аналітиком під час вирішення завдань.

У загальному випадку процес формування моделі користувача відбувається таким чином. Первинна модель користувача вважається множиною понять щодо досліджуваної предметної галузі, за якою аналітику необхідно знайти релевантну інформацію.

На підставі загальної моделі предметної галузі (моделі знань та вмінь технолога) відбувається постановка завдання. Аналіз результатів та процесу отримання цих результатів використовується для створення ситуаційної моделі (оцінка понять, включених для пошуку інформації за досліджуваною темою).

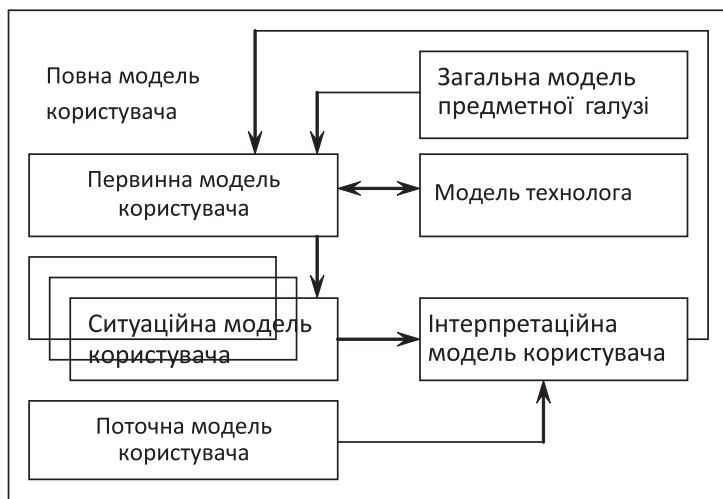


Рис.1. Модель користувача в ІС

Узагальнюючи окремі складові ситуаційної моделі, формуємо інтерпретаційну модель користувача. На наступному кроці інтерпретаційна модель користувача комбінується з поточною моделлю. Результатом є повна модель користувача, зображена на рис.1.

Модель користувача, яка використовується в перцепційно-інтерпретаційній технології [2], представлена графом, вузли якого відповідають поняттям, а дуги – взаємозв'язку між ними. Кожний вузол та дуга мають деяку величину або набір величин (допускається спадкування величин), які характеризують ступінь значущості даного поняття для аналітика.

Формування моделі користувача в ІС відбувається з огляду на тип інформаційного завдання, що розв'язується в системі (ознайомлення, конкретний пошук). Під час попереднього ознайомлення з досліджуваною предметною галузю (широкий погляд на проблему) модель користувача складається з множини тем для пошуку, що представлено онтологіями предметних галузей за напрямками роботи відповідних аналітичних підрозділів. При дослідженні предметної галузі аналітик має конкретне пошукове завдання, реалізація якого передбачає надходження відповідних документальних матеріалів або мультимедійної інформації, що враховується при формуванні первинної моделі користувача.

Таким чином, попередня модель користувача складається з теми для пошуку та множини понять з ваговими коефіцієнтами, які

визначають важливість (значущість) вказаного поняття для аналітика:

$M_k = \langle G', U' \rangle$, де G' – множина об'єктів моделі користувача, U' – множина дуг, що пов'язують об'єкти моделі користувача. Кожна дуга показує взаємний зв'язок між темами або співвідношення між темами понять, які використовуються аналітиком. Модель користувача має вигляд семантичної мережі, структура якої ізоморфна структурі семантичної мережі предметної галузі.

При зміні теми дослідження змінюється тема для пошуку, а відповідно, і складові моделі користувача. Зокрема, відбувається її розширення за рахунок множини понять теми G_i , при цьому змінюється значення показників, які визначають значущість понять, що входять до моделі користувача: $M_{k_{i+1}} = M_{k_i} \oplus G_i$, де M_{k_i} – модель користувача на попередньому кроці; G_i – об'єкт досліджуваної предметної галузі, обраний користувачем; \oplus – операція додавання вершини до семантичної мережі. Показники, що визначають значущість понять у M_{k_i} , розраховуються, як середнє значення релевантності поняття на попередньому кроці ϑ_{i-1} та релевантності поняття

інформаційної частини об'єкта δ_{G_i} : $\vartheta_i = (\vartheta_{i-1} + \delta_{G_i})/2$

Модель користувача складається з множини отриманих документів, тем, які використовувалися для пошуку, понять предметної галузі, що досліджується аналітиком та має вигляд семантичної мережі. На рис. 2 зображено структуру, що утворюється внаслідок встановлення зв'язків U_i між елементами, представленими темами пошуку, документами семантичної мережі предметної галузі й складовими моделі користувача.

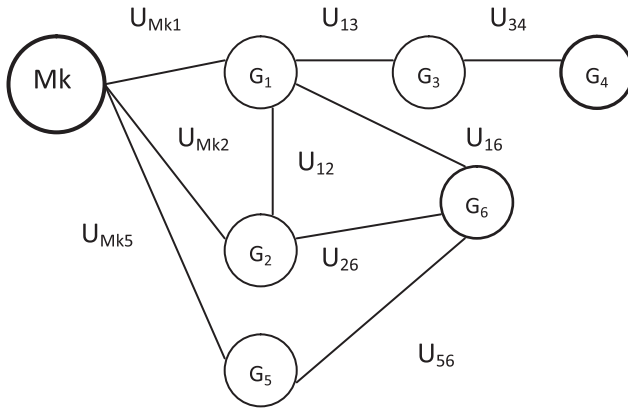


Рис. 2. Семантична мережа з елементами моделі користувача

У разі розв'язання інформаційного завдання, спрямованого на ознайомлення аналітика з досліджуваною предметною галузю, відбувається формування словника понять, встановлення відповідних вагових коефіцієнтів та тем для пошуку. При перегляді отриманих даних аналітик визначає пертинентну інформацію. На основі цих даних система здійснює адаптацію моделі користувача відповідно до його інформаційної потреби.

Підвищення рівня інтелектуальності інформаційної системи потребує врахування в моделі користувача дій аналітика, які йому необхідно здійснити для задоволення інформаційної потреби. Модель вказаних дій має вигляд упорядкованої послідовності: m_1, m_2, \dots, m_n , де кожний елемент m_i представлено моделлю користувача, на стан якої впливає тема пошуку та релевантність отриманих результатів.

Розглянемо нижче випадок, коли аналітик розв'язує інформаційне завдання, пов'язане з пошуком даних стосовно досліджуваної предметної галузі. Зокрема, модель дій користувача має деревоподібну структуру й містить дві складові. Перша – теми, що були обрані та ігноровані системою під час пошуку. Друга – документи, обрані, ігноровані та отримані згідно з темою пошуку й вимогами аналітика. За результатами аналізу моделі користувача та семантичної мережі в інтелектуальній інформаційній системі відбувається формування варіантів можливого вибору аналітика (рис. 3).

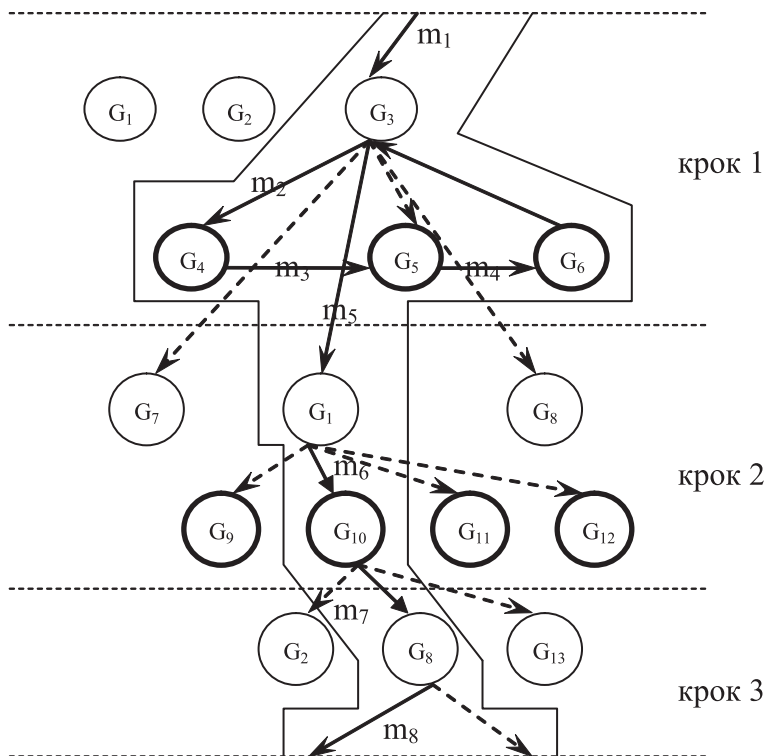


Рис. 3. Модель дії користувача в ІС

У наведеній вище структурі ребро (m_1) означає, що для дослідження предметної галузі аналітиком обрано тему G_3 , системою опрацьовано запит та надано відповідні документи. Ребра (m_2, m_3, m_4) – аналітик вивчає документи G_4, G_5, G_6 , проте не обирає серед них pertinentних, з урахуванням цього система формує для аналітика множину тем схожих з темою G_3 для можливого подальшого пошуку. Ребра (m_5, m_6) – користувач для пошуку обирає тему G_1 , система виконує запит та надає множину документів, серед яких аналітик обирає G_{10} ; водночас система пропонує додаткову множину тем, близьких до G_{10} . Ребро (m_7) – для подальшого пошуку аналітиком обирається тема G_8 ; вказаний процес повторюється до задоволення аналітиком своєї інформаційної потреби.

Узагальнюючи вищенаведене, маємо: на першому кроці аналітик обирає тему для пошуку. На її основі формується множина документів, які надаються аналітику для опрацювання. На наступному кроці

обрані для перегляду матеріали враховуються системою, у результаті чого змінюється модель користувача, і на її основі створюється нова множина документів, яку отримує аналітик для перегляду. У подальшому система аналізує документи, які зацікавили аналітика та синтезує наступні теми для пошуку, які за своїм контекстом релевантні темам, що вивчалися раніше. Також можливий інший варіант – коли система надає для пошуку нову тему з огляду на інтерес аналітика до конкретного документа (алгоритмом враховується список понять з даного документа).

На наступному етапі буде розглянуто процес ознайомлення аналітика з досліджуваною предметною галузю. У цьому випадку модель дій користувача представлено лінійною упорядкованою послідовністю m_1, m_2, \dots, m_n , якою визначається порядок надання тем користувачу, а також поняття, які належать до обраних тем.

Інтелектуальною інформаційною системою на основі моделі досліджуваної предметної галузі (представлена темами), аналізу семантичної мережі бази знань та з урахуванням результатів опитування аналітика здійснюється формування множини тем з подальшим ознайомленням користувача.

Визначивши досліджувану предметну галузь та множину понять, які необхідні аналітику для пошуку, система визначає тему (множину тем), для якої вказані поняття вважаються вихідними (тобто документи, знайдені за даним запитом, повинні визначати ці поняття). У подальшому, для обраних тем визначаються вхідні поняття. Системою реалізується зворотний зв'язок, суть якого полягає в отриманні від аналітика результатів оцінки цих понять.

Водночас для тих понять, з якими виникає ускладнення в оцінці, система намагається віднайти теми, де такі поняття будуть вихідними. У разі потреби, із бази знань користувачу буде надано документи, які мають максимальну вагу цих понять. Вказана операція виконується до визначення всіх тем для пошуку, а також понять, які входять до них.

З метою визначення оціночної функції нижче буде розглянуто процес використання моделі користувача для пошуку інформації.

Так, інтелектуальною інформаційною системою при задоволенні інформаційної потреби аналітика використовується критеріально-екстремізаційний механізм вибору. Згідно з запитом користувача проводиться формування функції пріоритетів.

Під час пошуку інформації, необхідної аналітику, існує декілька варіантів визначення опорних понять, які використовуються з метою ранжирування документів, зокрема:

– аналітик тільки почав працювати, модель користувача не сформована, не задана множина понять, яка визначає мету пошуку. У такому разі система формує первинну модель користувача, яка містить множину всіх понять, що існують в інтелектуальній інформаційній системі з нульовим коефіцієнтом значущості (у аналітика відсутній запит на пошук);

– аналітик почав працювати й вибрав тему для пошуку з досліджуваної предметної галузі, а також визначив множину понять з даної теми. При цьому формується первинна модель користувача, що містить множину даних понять з коефіцієнтами значущості заданими користувачем, і тему для пошуку з множини тем предметної галузі, які існують у системі;

– аналітик працює з інтелектуальною інформаційною системою. У цьому випадку система використовує множину понять, визначену на множині обраних тем, які містяться в моделі користувача.

Відобразимо на площині Ψ релевантність документів, наданих системою за результатом виконання пошукового запиту. Визначимо на площині місце окремих документів. Вказана площина показує рівень, який бажано досягти з метою отримання документів, здатних задовольнити інформаційну потребу аналітика.

Оцінка $\Psi_{\text{МК}}(G_k)$ визначає релевантність знайденого документа відповідно до запита аналітика:
$$\Psi_{\text{МК}}(G_k) = \sum_{j=1}^J \vartheta_j \sum_{i=1}^I \vartheta_i \delta_s(H_i, H_j)$$

де J – кількість понять, які належать до моделі користувача; I – кількість понять, які належать документу G_k ; ϑ_j – значущість поняття в моделі користувача; ϑ_i – значущість поняття в документі; $\delta_s(H_i, H_j)$ – релевантність i -го поняття J -му.

Зокрема, зростання значення $\Psi_{\text{МК}}(G_k)$ свідчить про збільшення пертинентності знайдених документів. Чисельне значення $\Psi_{\text{МК}}(G_k)$, яке відповідає документу G_k , вважатимемо критеріальною оцінкою, а отриману внаслідок цього шкалу – критеріальною. Таким чином, документи, пошук яких відбувається за запитом аналітика, утворюють

множину, що задовольняє такій умові: $\max_{G_k \in G} \psi(G_k)$ – варіанти,

які мають максимальну критеріальну оцінку при порівнянні з усіма іншими варіантами документів, що належать до моделі досліджуваної предметної галузі.

Наведена вище функція оцінки $\psi_{Mk}(G_k)$ може використовуватись як при звичайному, так і при розширеному способі пошуку інформації.

Проведення розширеного пошуку інформації передбачає використання коефіцієнтів для розрахунку функції оцінки. Зокрема, для цільового списку понять встановлюється коефіцієнт β_r , задається користувачем та міститься в межах від 0 до 1.

Функція оцінки в загальному випадку має такий вигляд:

$$\psi_{Mk}(G_k) = \sum_{j=1}^J \left(\beta_j \vartheta_j \sum_{i=1}^I (\beta_i \vartheta_i \delta_s (H_i, H_j)) \right)$$

у разі коли «i» та «j» поняття належать цільовому списку, використовуємо β , у іншому – $\beta^* = 1 - \beta$.

З наведеного видно, що збільшення важливості понять із цільового списку понять відбувається, якщо $\beta > \beta^*$, у даному випадку

знайдені документи стають більш значущими. При $\beta < \beta^*$ відбувається зменшення важливості понять із цільового списку понять, і знайдені документи стають менш значущими. При $\beta = 1$ маємо звичайний пошук, $\beta = 0$ – максимально розширений пошук з використанням усіх понять досліджуваної предметної галузі. За допомогою даного способу можливо шукати документи, які близькі до понять із ключової теми та переглядати документи, які мають побічний зв'язок з поняттями, що визначають ціль пошуку.

Водночас при проведенні розширеного пошуку аналітику можуть надаватися документи згідно з ключовою темою пошуку, а також документи, які були знайдені відповідно до тем пошуку, що містяться на нижчих рівнях ієрархії відносно ключової теми та належать до досліджуваної предметної галузі (рис. 4).

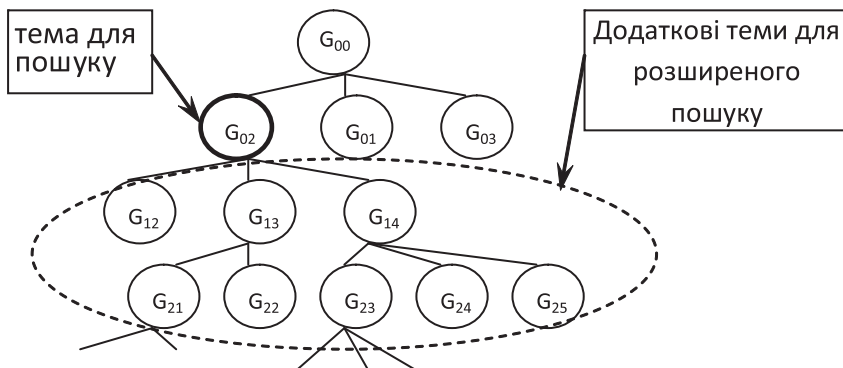


Рис.4. Додаткові теми для розширеного пошуку

При проведенні звичайного пошуку передбачається розгляд понять, які належать тільки до списку понять із ключової теми пошуку. Документи, що не містять понять із ключової теми, виключаються з подальшого розгляду, а функції оцінки обчислюються так само, як і при розширеному пошуку.

Зокрема, використання понять, заданих ключовою темою пошуку, дає можливість, по-перше, знаходити тільки ті документи, що пов'язані з цими поняттями, по-друге, проводити візуалізацію отриманих результатів за релевантністю (на основі моделі користувача). Процес візуалізації передбачає формування наочної моделі, яка дає змогу аналітику провести комплексну оцінку знайдених інформаційною системою документів з огляду на їх значущість для дослідження предметної галузі.

Для якісного проведення групування документів необхідно визначити такі властивості тем для пошуку [3]:

- *цільність теми* – характеризує інтенсивність розташування документів у просторі моделі предметної галузі;
- *дисперсія теми* – характеризує ступінь розсіювання документів у просторі відносно теми і показує, наскільки близько один до одного розташовані в семантичній мережі документи;
- *розмір теми* – визначає кількість документів, що належать до теми.

Формування масиву документів за темою пошуку відбувається на основі аналізу множини документів з досліджуваної предметної галузі. Алгоритм відбору документів із простору семантичної мережі пошукової системи складається з таких етапів: розподіл моделі предметної

галузі на теми для пошуку; створення груп документів, близьких до заданої теми (використовується функція оцінки); порівняння кожного документа з утвореними групами, центром яких є тема для пошуку.

Як характеристика відхилення документа в групі використовується середня сума квадратів відхилень від теми пошуку:

$$\Delta_s = \frac{1}{J} \sum_{j=1}^J \left(\frac{1}{N_j} \sum_{i=1}^{N_j-1} (\psi_n(D_i) - 1)^2 \right)$$

де Δ_s – середнє відхилення документів у групах від теми пошуку, J – кількість тем пошуку в моделі предметної галузі, N_j – кількість

документів у j -й темі, $\psi_n(D_i)$ – нормалізована функція оцінки близькості документів до теми пошуку:

$$\psi_n(D_i) = \frac{1}{P_t} \sum_j^P \vartheta_j \frac{1}{P_d} \sum_{i=1}^{P_d} \vartheta_i \delta_s(H_i, H_j)$$

де P_t – кількість понять у темі пошуку,

P_d – кількість понять у документі D_i .

Після визначення теми відбувається первинна вибірка документів, які надаються аналітику для подальшого опрацювання. У результаті вибору користувача визначається належність знайденого документа до теми пошуку, і за допомогою функції оцінки з групи документів відбираються остаточні інформаційні матеріали.

Аналітик визначає для себе значущість знайденого документа. Вказане враховується системою, унаслідок чого модель користувача адаптується до його потреб. Зокрема, позитивна відповідь збільшує значущість понять у моделі користувача, що належать знайденому документу, а негативна, відповідно, зменшує. Кількісна оцінка значущості розраховується за такою формулою:

$$\vartheta = \frac{1}{2} (\vartheta^* + \beta \delta_D(P, D_i)),$$

де ϑ^* – значущість i -го поняття для користувача, $\delta_D(P, D_i)$ – близькість поняття до знайденого документа, β – коефіцієнт, обирається залежно від ставлення користувача до знайденого документа: 1 – при задоволенні документом, 0 – у іншому випадку.

З огляду на викладене вбачається, що проведення адаптації моделі користувача відповідно до його інформаційної потреби дасть змогу

інтелектуальній інформаційній системі якісно відпрацювати пошукові запити аналітика, а також знайти в базі знань множину найбільш релевантних для досліджуваної предметної галузі документів.

Список використаних джерел

1. *Валькман, Ю. Р.* Интеллектуальные технологии исследовательского проектирования: формальные системы и семиотические модели [Текст] / Ю. Р. Валькман. – К. : Port-Royal, 1998. – 250 с.

2. *Гнусов, А. А.* Информационно-поисковые системы на основе мультиагент-ного подхода [Текст] / А. А. Гнусов // Материалы V науч.-практ. семинара «Новые информационные технологии». – М.: МГИЭМ, 2002. – 242 с.

3. *Дюран, Б., Оделл П.* Кластерный анализ [Текст] / Б. Дюран, П. Оделл. – М. : Статистика, 1977. – 128 с.

4. *Дриянский, В. М.* Модель поиска в интерактивных документальных информационных системах: аналитический обзор [Текст] / В. М. Дриянский // Кибернетика. – 1981. – № 2. – С. 102–118.

5. *Козлов, Е. Б., Хорошевский, В. Ф.* Об одной модели оценки релевантности результатов поиска информации на основе онтологии [Электронный ресурс] /Е. Б. Козлов, В. Ф. Хорошевский // Материалы VIII нац. конф. по искусственному интеллекту «КИИ-2002». – Режим доступа: <http://www.dialog-21.ru>, свободный. Заглавие с экрана.