ЗМІСТ

|  |  |
| --- | --- |
| Перелік умовних позначень, символів, одиниць, скорочень і термінів ……. | 999 |
| Вступ ….…..………………………………………………………………...… | 999 |
| 1 Огляд і аналіз сучасного стану проблеми автоматизації виробничо-збутових процесів ...................................................................................…….… | 999 |
| 1.1 Структура виробничо-збутової системи …………………….………… | 999 |
| 1.2 Пошукові процедури в сфері збуту товарів …………………………… | 999 |
| 1.2 Виробничо-збутові процеси як об'єкти управління. Модель Форрестера …………………………………………………………..………… | 999 |
| 1.4 Управління виробничо-збутовими процесами. Функції пошуку в виробничо-збутових системах …………………..………………………….... | 999 |
| 1.5  Постановка задачі …………………………………………………...… | 999 |
| 2 Методи пошуку в інформаційних системах ………...……………...……… | 999 |
| 2.1 Інформаційний пошук як процес методи інформаційного пошуку …………..…………..…………..…………..…………..……………… | 999 |
| 2.2 Організація пошуку. Стратегія пошуку ……………………….……… | 999 |
| 2.3 Пошукові методи в інформаційних системах. Нечіткий пошук …………………………………………………………………………… | 999 |
| 2.3.1 Відстань Левенштейна. Метод Вагнера-Фішера ………..………. | 999 |
| 2.3.2 Алгоритми нечіткого пошуку без індексації ……………………. | 999 |
| 2.3.2.1 Лінійний пошук ……………….……………….……………… | 999 |
| 2.3.2.2 Метод Bitap з модифікаціями від Wu і Manber …………….. | 999 |
| 2.3.3 Алгоритми нечіткого пошуку з індексацією …………………….. | 999 |
| 2.3.3.1 Метод N-грам ….…..………………………………….……… | 999 |
| 2.3.3.2 Метод розширення вибірки ………………………………….. | 999 |
| 2.3.3.3 Хешування за сигнатурою …………………………………… | 999 |
| 2.3.3.4 BK-дерева ……………………………………………………… | 999 |
| 2.4 Аналіз недоліків та варіанти модифікації існуючих алгоритмів ..…. | 999 |
| 3 Дослідження і програмна реалізація пошукових методів. Оцінка  ефективності пошукових методів ………………………………………......... | 999 |
| 3.1 Вибір середовища і технологій програмування ….…..……………… | 999 |
| 3.2 Проектування та реалізація програми для оцінки ефективності пошукових методів ….…..………………….…..………………….…..……… | 999 |
| 3.3 Аналіз результатів програми .…..…………………..…..…………….. | 999 |
| Висновки ….…..……………………………………………………………….... | 999 |
| Перелік посилань …………………………………………….…………..……. | 999 |

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,  
СКОРОЧЕНЬ І ТЕРМІНІВ

|  |  |
| --- | --- |
|  |  |
| ВЗС | –    Виробничо-збутова система |
|  |  |
| ІП | –    Інформаційний пошук |
| ІПС | –    Інформаційна пошукова підсистема |
| ПЗ | –    Пошуковий запит |

ВСТУП

Системи управління виробництвом і збутом продукції підприємства розвиваються з моменту їх зародження та перетерпіли зміни різних форм організації: від системи ручного збору інформації про запаси і потужності до сучасних автоматизованих систем контролю та обліку, що працюють на базі новітніх ЕОМ, які застосовуються від найменших і до найбільших підприємств. Сучасні системи управління дозволяють контролювати весь процес виробництва і збуту продукції в умовах постійної зміни зовнішнього і внутрішнього середовища підприємства [1].

Однак останнім часом постійно збільшується число задач, для вирішення яких виявляється доцільно та необхідно створювати нові сучасні системи управління. В умовах сталої тенденції до збільшення складності і різноманітності задач управління, існуючі підходи до побудови систем управління вже не здатні забезпечити необхідний рівень управління та адаптації. Аналіз поточного стану в області розробки автономних адаптивних систем управління дозволяє виділити наступні основні проблеми, з якими стикаються розробники подібних систем [2]:

– проблема координації різних форм поведінки і організації взаємодії між ними;

– проблема планування дій: система управління повинна не просто реагувати на поточну ситуацію, а й прогнозувати послідовність дій, які повинні привести до наміченої мети;

– проблема навчання на досвіді взаємодії з зовнішнім середовищем; дана проблема включає не тільки питання навчання окремої формі поведінки, але питання навчання взаємодії між різними формами поведінки;

– проблема формування нових типів поведінки: система управління повинна вміти самостійно формувати нові типи поведінки, спрямовані на досягнення нових цілей, і включати їх в загальну структуру контролю;

– проблема універсальності: системи управління різними об'єктами повинні слідувати єдиним принципам управління і базуватися на єдиній моделі.

Існує безліч складних задач, які необхідно вирішити підприємству, для свого успішного функціонування. До таких задач відносяться [1]:

– скорочення обсягів сировини і незавершеного виробництва;

– скорочення періоду виробничого циклу;

– зниження собівартості продукції за рахунок поліпшення організації виробничого процесу;

– покращення задоволення попиту.

Зазвичай отримання додаткового прибутку пов'язане зі зміною або розширенням діяльності підприємства на ринку, які можливі за збільшення обсягу продажів наявних послуг на існуючих ринках за рахунок більш активної маркетингової політики [3].

Прибутковість фірми свідчить про ефективність її виробничо-збутової діяльності, яка досягається шляхом мінімізації витрат (на сировину, матеріали, енергію, оплату праці, фінансування) і максимізації доходів від результатів виробництва – випуску продукції і послуг. Для досягнення більшої ефективності та доходів компанії, слід подбати про автоматизацію процесів пошуку, вибору та купівлі товарів клієнтом.

Метою роботи є дослідження пошукових процедур в системах керування виробничо-збутовими процесами, оцінка ефективності пошукових алгоритмів та методів пошуку інформації, визначити методу, який буде мати максимальну точність, повноту та швидкість (продуктивність) пошуку, а також здійснювати коригування пошукових запитів, використовувати нечітку логіку знаходження і видачі результатів, що сприятиме підвищенню ефективності виробничо-збутової системи.

1 ОГЛЯД І АНАЛІЗ СУЧАСНОГО СТАНУ ПРОБЛЕМИ АВТОМАТИЗАЦІЇ ВИРОБНИЧО-ЗБУТОВИХ ПРОЦЕСІВ

1.1 Структура виробничо-збутової системи

Як відомо, головною метою капіталу в будь-якій сфері діяльності є висока ефективність виробництва, яка відноситься до числа ключових категорій ринкової економіки і безпосередньо пов'язана з досягненням кінцевої мети розвитку суспільного виробництва в цілому, а також кожного підприємства окремо. У загальному вигляді економічна ефективність являє собою кількісне співвідношення двох величин – результатів діяльності та виробничих витрат. Сутність проблеми підвищення економічної ефективності виробництва полягає в збільшенні економічних результатів на кожну одиницю витрат в процесі використання наявних ресурсів [3].

Управління виробничо-збутовими процесами є досить трудомісткою діяльністю, що включає в себе облік замовлень, продажів, попиту і потреб клієнтів, а також аналіз отриманих даних та прийняття на їх основі правильних управлінських рішень.

При розгляді збутових процесів окремо, можна визначити, що вони починаються з того моменту, як виріб покинув виробничий комплекс, і закінчуються передачею товару покупцеві. Це являє процес реалізації виробленої продукції з метою перетворення товарів у гроші й задоволення запитів і потреб покупців [4]. Процес управління збутом представлений на рисунку 1.1.

Таким чином, сутність управління збутовою діяльністю розкривається в таких основних управлінських функціях, як планування, організація, мотивація та контроль. Об'єктами цих функцій виступають, наприклад, ресурси, задіяні в процесі збуту, персонал, клієнти компанії, канали просування продукції, організаційні форми управління продажами.

Існуючі на сьогоднішній день підходи до вирішення управлінських проблем підприємств є недостатньо ефективними. Це пов’язано з [5]:

– ускладненням процесів прогнозування попиту;

– складністю формалізованого уявлення управлінських ситуацій в інформаційних системах;

– наявністю постійно змінної специфіки взаємовідносин з діловими партнерами в частині управління ціноутворенням і наданням знижок і бонусів;

– відсутністю концептуальних положень застосування інструментальних засобів, що враховують галузеву специфіку.

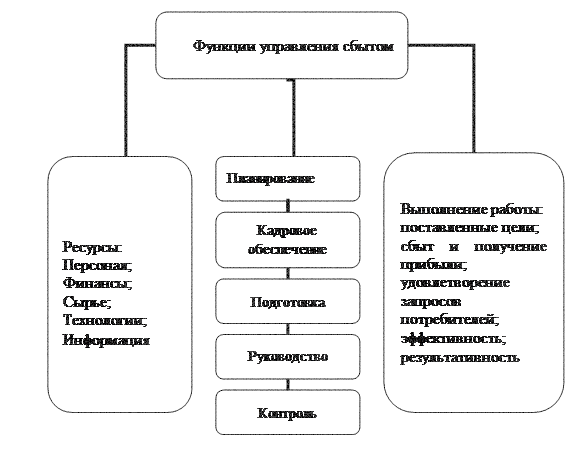


Рисунок 2.1 – Процес управління збутом

Отже, яким чином організаційна форма й правила прийняття рішень можуть стати джерелом типових небажаних явищ у поводженні об’єкта в цілому. Найважливіша проблема виробничо-збутових системи є приведення темпу виробництва й темпів продажу продукції у відповідність до вимог кінцевого споживача [6].

Як показує практика, темпи виробництва часто змінюються в більших межах, ніж фактичні темпи споживчих покупок.

Виявлено, що виробничо-збутові системи (ВЗС) з ланцюгами взаємозалежних товарних запасів і певним порядком видачі замовлень на їх поповнення мають властивості підсилювати невеликі коливання, що виникають у роздрібній ланці.

Глобальною метою моделювання подібних об’єктів є пошук відповідей, на такі питання [6]:

– як невеликі зміни обсягу роздрібних продаж можуть викликати значні коливання виробництва продукції підприємства?

– чому прискорення виконання конторських робіт може не зробити істотного впливу на поліпшення управлінських рішень?

– чому керівництво підприємством може виявитися не в змозі виконати замовлення, хоча його виробничі можливості перевищують обсяги продаж?

Об’єктом дослідження є трирівнева територіально розподілена система з виробництва і збуту продукції, що включає завод (виробнича ланка), мережу оптових баз (оптова ланка) і мережу магазинів (роздрібна ланка; рис. 1.2).

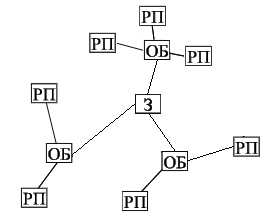


Рисунок 2.2 – Структура виробничо-збутової системи:  
З – завод; ОБ – оптова база; РП – роздрібне підприємство

У системі існують запаси трьох рівнів: на заводі, в оптовій і у роздрібній ланках. Запізнювання, що мають місце у системі подано у тижнях (рис. 1.3), являють собою звичайні величини для підприємств, що виготовляють товари довготривалого користування. Передбачено три види замовлень: замовлення на відшкодування проданих товарів; замовлення для поповнення запасів у всіх ланках у зв'язку зі зміною рівня продаж; замовлення, необхідні для заповнення каналів забезпечення товарами по замовленнях, що знаходяться на стадії виконання. У системі прийнято такий порядок видачі замовлень. На основі аналізу продажів і, у відповідності з запізнюванням закупівлі, замовлення найближчій вищій ланці системи включають відшкодування фактичних продажів. Після проходження достатнього часу для визначення середньої величини короткострокових коливань продажів (вісім тижнів) приймаються заходи для поступового зниження (або підвищення запасів залежно від збільшення або зменшення попиту на товар). Обсяг замовлень, що перебувають у процесі виконання (відправлені поштою, невиконані замовлення у постачальника, товари на стадії доставки) прийнятий пропорційним середньому рівню ділової активності і тривалості виконання замовлення.



Рисунок 1.3 – Організаційна структура ВЗС

Із використанням пошукової підсистеми та підсистеми віддаленого збуту товарів, ВЗС надасть можливість покупцям робити пошук та покупки товару віддалено. Додавши аналітичні методи, на основі аналізу продажів та пошукових запитів можна з більшою точністю визначати і прогнозувати необхідність виготовлення продукції, закупівлю необхідних матеріалів тощо.

Інформаційна пошукова підсистема (ІПС) є частиною автоматизованої системи управління виробничо-збутовими процесами та виконує наступні задачі:

– автоматизація процесів пошуку товарів;

– збір та обробка статистики пошукових запитів.

Використання інформаційної пошукової підсистеми для виробничо-збутової системи сприяє вирішенню таких завдань:

– скорочення обсягів сировини і незавершеного виробництва;

– скорочення періоду виробничого циклу;

– поліпшення задоволення попиту.

1.2 Пошукові процедури в сфері збуту товарів

Процес продажу товару проводиться таким чином. Клієнт (покупець) виконує замовлення товару на роздрібному підприємстві і, в разі наявності товару, отримує своє замовлення. В даному випадку, процес не займає багато часу, але давайте розглянемо той випадок, якщо товар (або деякі його комплектуючі) відсутні на роздрібному підприємстві. Необхідно провести пошук товару (комплектуючих) і запит на їх поставку на найближчу оптову базу або склад. Якщо товар не знайдений і на складі, запит проводиться на рівень вище, доки не досягне заводу, на якому товар вже буде виготовлений і в подальшому доставлений на роздрібне підприємство, з якого надійшло замовлення.

Таким чином, процес збуту займає досить багато часу, особливо при ВЗС складної структури (декількома заводами; рівнями складів і оптових баз і т. д.).

Для збільшення ефективності і скорочення термінів доставки товару клієнту, варто включити автоматизовану пошукову систему. Пошукова система дасть можливість за короткий термін виконувати процеси пошуку товару на складі або оптовій базі, а також аналізувати пошукові запити. На основі цього прогнозувати необхідність виготовлення і доставки продукції, закупівлю матеріалів [7].

Також дуже важливим є те, що товар або послуга повинні задовольняти потреби клієнтів. Виробничо-збутовому підприємству слід подумати про процеси пропозиції послуги, тобто про те, як саме варто пропонувати її клієнтам, і як робити це найбільш ефективно. Важливу роль відіграє процеси пошуку товарів клієнтами.

Варіант віддаленого пошуку і продажу товарів є досить актуальним на сьогоднішній день, так як сьогодні мільйони людей щодня, не виходячи з дому, купують різні товари в електронних магазинах. У світі, а зокрема в Україні, величезними темпами зростає кількість користувачів Internet і, як наслідок, кількість «електронних» покупців і потенційних «електронних» покупців [8].

Електронні магазини істотно зменшують витрати виробника, заощадивши на утриманні звичайного магазину, а також розширюють ринки збуту, так само як і розширює можливість покупця – купувати будь-який товар в будь-який час та в будь-якому місці. Це надає електронним магазинам безперечні переваги перед звичайними магазинами. Даний факт є істотним при переході виробників від «звичайної» торгівлі на «електронну» [8].

Висока якість продукції, вміння донести інформацію про продукт до споживача і ефективна система збуту, робить підприємство успішним на ринку. У багатьох компаніях зустрічаються проблеми збуту, які заважають ефективно працювати відділу продажів, і не зникають навіть з підбором хороших продавців [8]. Вирішити їх можна лише шляхом автоматизації процесу продажів. В даних системах важливими є безпосередньо послуги, програмні можливості та інструменти, що надаються клієнтам для пошуку товарів. Загальний принцип роботи систем автоматизації процесів пошуку товарів полягає в тому, що для пошуку інформації користувач (клієнт) формулює запит [9]. Функцією пошукової системи є безпосередньо знайти за запитом користувача та надати товари (послуги), що містять зазначені ключові слова або назву, чи слова, будь-яким чином пов’язані з ключовими словами запиту клієнта [10].

1.3 Виробничо-збутові процеси як об'єкти управління. Модель  
Форрестера

З метою розглянути виробничо-збутову систему, як об'єкт управління, необхідно отримати модель системи. При позначенні змінних і констант використаються буквені ідентифікатори, останні символи яких відповідають першим літерам назви ланки:

– R – роздрібна;

– O – оптова;

– P – виробництва.

Складання моделі зручно почати із двох простих рівнянь що описують обсяги невиконаних замовлень nzr і запасів товарів на складі fzr [xx]:

*(1–u): nzr.k = nzr.j + dt\*(pzr.jk – otr.jk);* (1.1)

*(2–u): fzr.k = fzr.j + dt\*(ptr.jk – otr.jk),* (1.2)

де *nzr* – рівень невиконаних замовлень у роздрібній торгівлі, (од.);

*dt* – крок моделювання (тижд.);

*pzr* – поточний темп потоку замовлень до РЛ (од./тижд.);

*otr* – темп потоку відвантаження товарів із РЛ (од./тижд.);

*fzr* –рівень фактичних запасів у РЛ (од.);

*ptr* – поточний темп, надходження товарів до РЛ (од./тижд.).

Рівняння темпів відображають механізм рішень, що властивий системі. Вони повинні залишатися справедливими й досить точними при будь-яких навіть самих великих варіаціях змінних, які можуть мати місце в системі.

Під темпом відвантаження товарів покупцям otr розуміється об'єктивно обумовлений темп. Темп відвантаження повинен залежати від величини заборгованості по невиконаних замовленнях nzr і від рівня фактичних запасів fzr:

*(3-v): pоtr.k = nzr.k/zzr.k,* (1.3)

*(4-v): dotr.k = fzr.k/dt.*  (1.4)

*(5-t): *  (1.5)

де pоtr – допоміжна змінна, передбачуваний темп відвантаження товарів, (од./тижд.);

dotr – граничний (допустимий) темп відвантаження, при якому за час між рішеннями dt буде повністю витрачений товарний запас fzr;

otr – фактичний темп відвантаження товарів, (од./тижд.).

Якщо розглядати один вид товару на одному складі, то замовлення можуть виконуватися доти, поки не закінчаться товари із запасу. Можливість виконання замовлень падає до нуля, як тільки виснажуються запаси товарів. У більш загальному випадку матимемо: один вид товару – багато складів; багато видів товарів – один склад; багато товарів – багато складів. У кожному із цих випадків запаси одних товарів будуть вичерпані раніше інших і загальна можливість виконувати замовлення буде поступово знижуватися зі зниженням загального обсягу запасів.

З урахуванням бажаного (раціонального) рівня запасів запізнювання виконання замовлень можна визначати за співвідношенням:

*(6-v): zzr.k = mzr + uzr\*jzr.k/fzr.k,* (1.6)

де *zzr –* запізнювання виконання замовлень роздрібною ланкою, (тижд.);

*mzr –* мінімально можливе запізнювання виконання замовлень РЛ, (тижд.);

*uzr –* середнє запізнювання виконання замовлень РЛ, пов'язане з відсутністю на складі деяких видів товарів при загальному «нормальному» обсязі запасів, (тижд.);

*jzr* – бажаний запас у РЛ, (од.);

*fzr* – фактичний запас у РЛ, (од.).

Бажаний рівень запасів може бути прийнятий прямо пропорційним середньому рівню продаж:

*(7-v): jzr.k = kpr\*upr.k,* (4.7)

де *jzr –* бажаний рівень запасів у РЛ, (од.)

*upr* – середній темп надходження замовлень до РЛ, (од./тижд.);

*kpr –* коефіцієнт запасу. Постійна *kpr* являє собою число тижнів, протягом яких середній темп продажів може бути забезпечений за рахунок бажаного запасу *jzr*.

Середній темп продажів можна визначити, використовуючи показникове усереднення 1-го порядку:

*(8-u): upr.k = upr.j + dt\*(1/kur)\*(pzr.jk – upr.j),* (4.8)

де *upr –* середній темп надходження замовлень до РЛ, (од./тижд.);

*kur* – постійна часу усереднення вимог до РЛ, (тижд.);

*pzr* – поточний темп надходження замовлень до РЛ, (од./тижд.).

Розглянемо реакцію системи на +10% ступінчасту зміну темпу надходження замовлень pzr, що має місце у січні місяці (рис. 1.4). Реакція системи має коливальний повільно затухаючий характер. Виниклі коливання представлені темпами видачі замовлень, випуску продукції, розмірами запасів на складах й обсягом невиконаних замовлень.

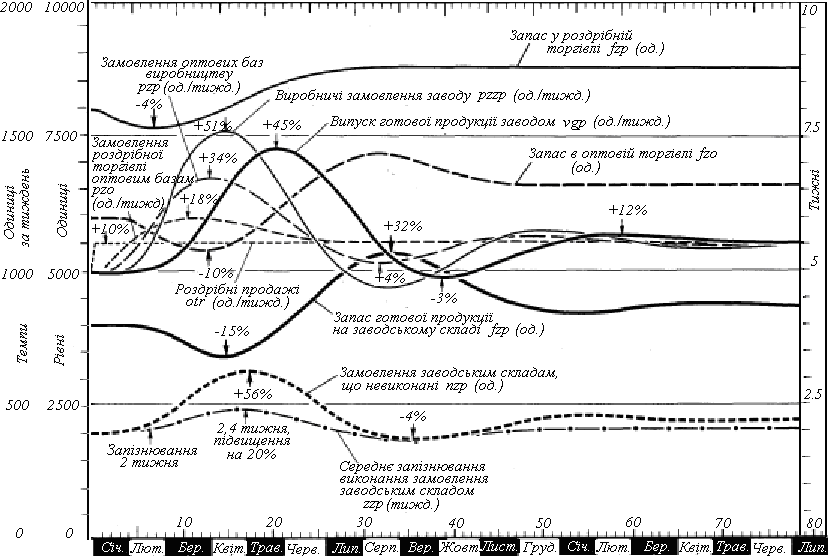


Рисунок 1.4 – Реакція системи на 10-процентне ступінчасте  
збільшення потоку замовлень

Припустимо, що в минулому система функціонувала в умовах постійного попиту на продукцію *pzr=1000*. Потім у січні продажі починають збільшуватися й наприкінці березня зростають на +10%, а до кінця вересня спостерігається 10-процентний спад і, нарешті, наприкінці грудня повернення до попереднього рівня *pzr=1000* [xx].

Поведінка системи у таких умовах є ще більш нестабільною (рис. 1.5). Початкове зростання темпів видачі замовлень і випуску продукції багато в чому схоже з розглянутим раніше. Відмінністю є те, що перші максимуми функцій дещо менші й відстають за часом. Однак у той час, коли система вже повинна була вийти зі стану пері виробництва, вона одержує додатковий понижуючий поштовх, викликаний спадом роздрібних продажів pzr. Як насідок, кількість замовлень, що надходять з ОЛ у ВЛ pzр, знизилась у жовтні в порівнянні зі звичайним рівнем на 40%, а обсяг випуску продукції vgp впав у листопаді на 60% у порівнянні з нормальним рівнем.

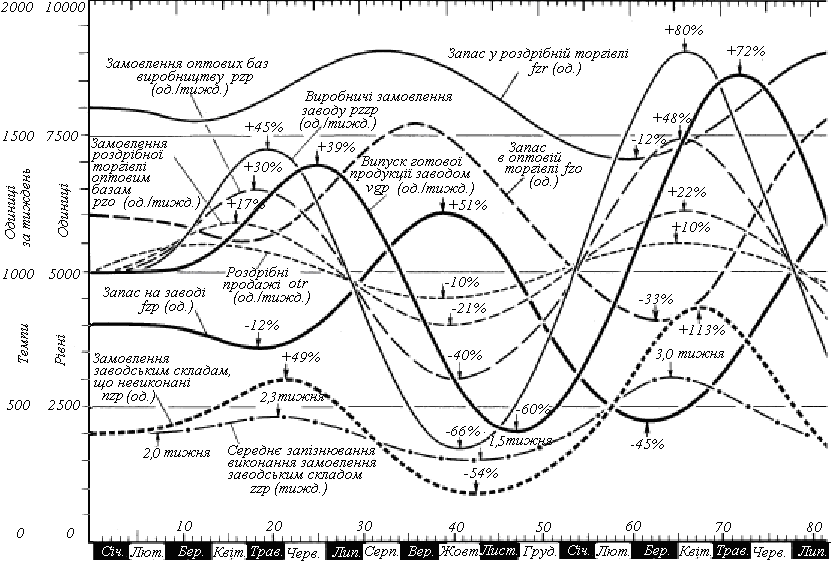


Рисунок 1.5 – Реакція системи на 10-процентне  
збільшення і падіння потоку замовлень

1.4 Управління виробничо-збутовими процесами. Функції пошуку в  
виробничо-збутових системах

1.5 Постановка задачі

Для виробничо-збутової системи необхідно на основі дослідження і аналізу пошукових процедур знайти та визначити оптимальну структуру реалізації інформаційної пошукової підсистеми, що сприяє рішенню наступних задач:

– скорочення обсягів сировини і незавершеного виробництва;

– скорочення періоду виробничого циклу;

– покращення задоволення попиту.

Необхідно провести дослідження та оцінку ефективності пошукових алгоритмів та методів пошуку інформації, та визначити такий метод, який буде мати максимальну точність, повноту та швидкість (продуктивність) пошуку, а також здійснювати коригування пошукових запитів, використовувати нечітку логіку знаходження і видачі результатів.

2 МЕТОДИ ПОШУКУ В ІНФОРМАЦІЙНИХ СИСТЕМАХ

2.1 Інформаційний пошук як процес. Методі інформаційного пошуку

Пошук інформації являє собою процес виявлення в деякій множині документів (текстів) всіх тих, які присвячені зазначеної теми (предмету), задовольняють заздалегідь визначеним умовам пошуку (запиту) або містять необхідні (відповідні інформаційної потреби) факти, відомості, дані.

Процес пошуку включає послідовність операцій, спрямованих на збір, обробку та надання інформації [11].

У загальному випадку пошук інформації складається з чотирьох етапів:

– визначення (уточнення) інформаційної потреби і формулювання інформаційного запиту;

– визначення сукупності можливих власників інформаційних масивів (джерел);

– вилучення інформації з виявлених інформаційних масивів;

– ознайомлення з отриманою інформацією і оцінка результатів пошуку.

Представимо процес інформаційного пошуку (ІП) даних в базі у вигляді моделі. Модель задачі інформаційного пошуку представляється у вигляді кортежу:

,         (2.1)

де, – масив даних (i = 1,n ; j = 1,m ; n – кількість таблиць в j-й базі даних; m – кількість баз даних);

– масив запиту;

– відношення ідентичності, тобто .

Пошук даних може здійснюватися різними методами.

*Адресний пошук.* Процес пошуку документів за формальними ознаками, зазначеним у запиті. Для здійснення потрібні такі умови [18]:

– наявність у документа точної адреси;

– забезпечення суворого порядку розташування документів в пристрої або в сховище систем.

Адресами документів можуть виступати адреси WEB-серверів і WEB-сторінок, а також елементи бібліографічного запису або адреси зберігання документів в базі даних.

*Семантичний пошук.* Процес пошуку документів за їх змістом [18].

Умови пошуку:

– переклад змісту документів і запитів з природної мови на інформаційно-пошукову мову і складання рейтингів образів документа і запиту;

– складання пошукового опису, в якому вказується додаткова умова пошуку.

Принципова різниця між адресним і семантичним пошуками полягає в тому, що при адресному пошуку документ розглядається як об'єкт з точки зору форми, а при семантичному пошуку – з точки зору змісту.

При семантичному пошуку знаходиться множина документів без вказівки адрес.

*Документальний пошук.* Процес пошуку в базі інформаційно-пошукової системи первинних документів або в базі даних вторинних документів, що відповідають запиту користувача [17, 18].

Існує два види документального пошуку:

– бібліотечний, спрямований на знаходження первинних документів;

– бібліографічний, спрямований на знаходження відомостей про документи, представлених у вигляді бібліографічних записів.

*Фактографічний пошук.* Процес пошуку фактів, відповідних інформаційним запитом. До фактографічних даних відносяться відомості, витягнуті з документів, як первинних, так і вторинних і одержувані безпосередньо з джерел їх виникнення [17, 18].

Розрізняють два види фактографічного пошуку:

– документально-фактографічний, полягає в пошуку в документах фрагментів тексту, що містять факти;

– фактологічний (опис фактів), що передбачає створення нових фактографічних описів в процесі пошуку шляхом логічної переробки знайденої фактографічної інформації.

2.2 Організація пошуку. Стратегія пошуку

Процес пошуку документів, релевантність (відповідність) щодо пошукового запиту, вивід результатів і т. п. Може бути реалізований по-різному. Даний процес називають стратегією пошуку. Стратегії інформаційного пошуку визначають ступінь подібності документів, що розглядаються, до пошукового запиту. Ступінь подібності визначається згідно з робочою гіпотезою: чим частіше пошуковий термін зустрічається в документі, тим «відповіднішим» є цей документ до пошукового запиту [11].

Стратегія інформаційного пошуку це алгоритм, який, переглядаючи набір документів (Д1, …, Дn), встановлює їх відповідність до пошукового запиту (ПЗ). Оскільки пошуковий термін зустрічається в документах різну кількість раз, можна говорити про різну ступінь відповідності до пошукового запиту.

Існують такі стратегії інформаційного пошуку [11]:

– з використанням векторно-просторового представлення (vector space model);

– пошук імовірності появи пошукового терміна в документі (probabilistic retrieval);

– з побудовою мовної моделі для кожного документа (language models);

– з побудовою мережі припущень, яка використовується для встановлення відповідності документу до пошукового запиту (inference network);

– з Булевим індексуванням, коли кожному пошуковому терміну присвоюється своя «вага», що потім враховується при побудові впорядкованих списків документів (Boolean indexing);

– з використанням не проявленого семантичного індексування (latent semantic indexing);

– з побудовою нейромереж (neural networks);

– з використанням продуктивних алгоритмів, коли початковий пошуковий запит «еволюційно» видозмінюється (genetic algorithms);

– з використанням нечітких множин, коли документу ставиться у відповідність нечітка множина (fuzzy set retrieval).

2.3 Пошукові алгоритми в інформаційних системах. Нечіткий пошук

Різні алгоритми реалізують різні стратегії. Виходячи с того, що розглядаються пошукові процедури в системи управління виробничо-збутих процесів, запит та дані в якій знаходяться в текстовому форматі, розглянемо переважно алгоритми нечіткого пошуку.

Нечіткий пошук є вкрай корисною функцією будь-якої пошукової системи. Разом з тим, його ефективна реалізація набагато складніше, ніж реалізація простого пошуку за точним збігом.

Задачу нечіткого пошуку можна сформулювати так: «По заданому запиту знайти в тексті або словнику розміру n все слова (записи), що збігаються (або містять) з цим запитом з урахуванням k можливих розходжень».

Алгоритми нечіткого пошуку характеризуються метрикою – функцією відстані між двома словами, що дозволяє оцінити ступінь їх подібності в даному контексті. Математичне визначення метрики включає в себе необхідність відповідності умові нерівності трикутника (X - безліч слів, p - метрика) [12]:

        (2.2)

У більшості випадків під метрикою мається на увазі більш загальне поняття, яке не потребує виконання такої умови, це поняття можна також назвати відстанню.

У числі найбільш відомих метрик - відстані Хеммінга, Левенштейна і Дамерау-Левенштейна. При цьому відстань Хеммінга є метрикою тільки на множині слів однакової довжини, що сильно обмежує область його застосування.

2.3.1 Відстань Левенштейна. Метод Вагнера-Фішера

Відстань Левенштейна (також редакційне відстань або дистанція редагування) між двома рядками має на увазі мінімальну кількість операцій вставки одного символу, видалення одного символу і заміни одного символу на інший, необхідних для перетворення одного рядка в іншу [18].

Редакційним приписом називається послідовність дій, необхідних для отримання з першого рядка другий найкоротшим чином. Зазвичай дії позначаються так: D (англ. Delete) - видалити, I (англ. Insert) - вставити, R (replace) - замінити, M (match) - збіг.

Наприклад, для 2-х рядків «CONNECT» і «CONEHEAD» можна побудувати таку таблицю перетворень (рис 2.1).

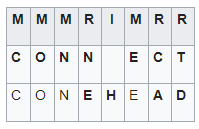


Рисунок 2.1 – Таблиця перетворень для двох рядків

*Різні ціни операцій.* Ціни операцій можуть залежати від виду операції (вставка, видалення, заміна) і / або від що беруть участь в ній символів, відображаючи різну ймовірність мутацій в біології [3], різну ймовірність різних помилок при введенні тексту і т. Д. У загальному випадку:

– w (a, b) - ціна заміни символу a на символ b;

– w (ε, b) - ціна вставки символу b;

– w (a, ε) - ціна видалення символу a.

Необхідно знайти послідовність замін, що мінімізувала сумарну ціну. Відстань Левенштейна є окремим випадком цієї задачі при:

– w (a, а) = 0;

– w (a, b) = 1 при a ≠ b;

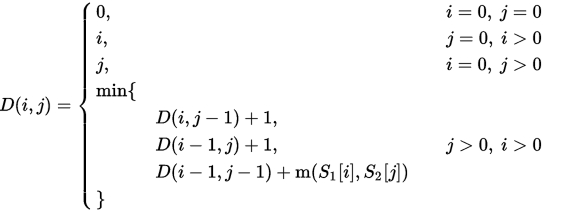
– w (ε, b) = 1;

– w (a, ε) = 1.

Як окремий випадок, так і завдання для довільних w, вирішує алгоритм Вагнера - Фішера, наведений нижче. Тут і нижче ми вважаємо, що все w невід'ємні, і діє правило трикутника: якщо дві послідовні операції можна замінити однією, це не погіршує загальну ціну (наприклад, замінити символ x на y, а потім з y на z не краще, ніж відразу x на z).

Якщо до списку дозволених операцій додати транспозицію (два сусідніх символу міняються місцями), виходить відстань Дамерау-Левенштейна. Для неї також існує алгоритм, що вимагає O (MN) операцій. Дамерау показав, що 80% помилок при наборі тексту людиною є транспозиція. Крім того, відстань Дамерау-Левенштейна використовується і в біоінформатики.

Нехай S1 і S2 – два рядки (довжиною M і N відповідно) над деякими алфавітом, тоді редакційне відстань (відстань Левенштейна) d (S1, S2) можна підрахувати за такою рекуррентной формулою

(2.3)

де m (a, b) дорівнює нулю, якщо a = b і одиниці в іншому випадку; min {a, b, c} повертає найменший з аргументів.

Тут крок за i символізує видалення (D) з першого рядка, по j - вставку (I) в перший рядок, а крок за обома індексами символізує заміну символу (R) або відсутність змін (M).

Початковий варіант цього методу має часову складність O (mn) і споживає O (mn) пам'яті, де m і n – довжини порівнюваних рядків. Весь процес можна представити матрицею, зображеною на рисунку 2.2 [19].

Якщо подивитися на процес роботи алгоритму, нескладно помітити, що на кожному етапі використовуються тільки дві останні рядки матриці, отже, споживання пам'яті можна зменшити (рис.2.3) до O(min (m, n)).

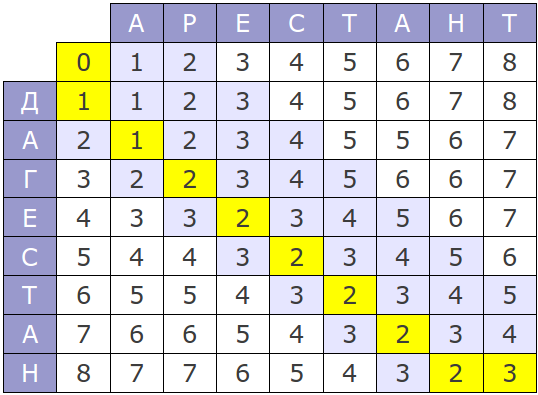


Рисунок 2.2 – Процес методу Дамерау-Левенштейна

Але це ще не все - можна далі оптимізувати алгоритм, якщо стоїть завдання знаходження не більше k відмінностей. В цьому випадку потрібно обчислювати в матриці лише діагональну лінію шириною 2k + 1 (відсікання Укконена), що зводить часову складність до O(k min (m, n)).

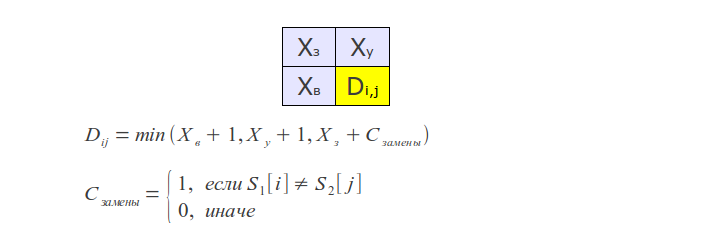


Рисунок 2.3 – Зменьшення споживання пам’яті

Метод у вигляді, описаному вище, вимагає Theta (M \* N) операцій і таку ж пам'ять. Останнє може бути неприємним: так, для порівняння файлів довжиною в 10 ^ 5 рядків буде потрібно близько 40 гігабайт пам'яті.

2.3.2 Алгоритми нечіткого пошуку без індексації

2.3.2.1 Лінійний пошук

2.3.2.2 Метод Bitap з модифікаціями від Wu і Manber

Двійковий алгоритм пошуку підрядка (також bitap algorithm, shift-or algorithm) використовує той факт, що в сучасних комп'ютерах бітовий зсув і побітовое АБО є атомарними операціями. По суті, це примітивний алгоритм пошуку з невеликою оптимізацією, завдяки якій за одну операцію проводиться до 32 порівнянь одночасно (або до 64, в залежності від розрядності машини) [20].

Обчислювальна складність - O (| needle | · | haystack |) операцій з вкрай малою константою. На попередню обробку йде O (| needle | · | Σ |) операцій та пам'яті, де Σ - алфавіт. Якщо ж довжина шаблону пошуку (в символах) не перевищує довжину машинного слова (в бітах), складність виходить O (| haystack |) і O (| needle | + | Σ |) відповідно.

Оригінальна версія алгоритму має справу тільки з замінами символів, і, фактично, обчислює відстань Хеммінга. Але трохи пізніше Sun Wu і Udi Manber запропонували модифікацію цього алгоритму для обчислення відстані Левенштейна, тобто привнесли підтримку вставок і вилучень, і розробили на його основі першу версію утиліти agrep [19].

Операції в методі виконуються наступним чином (рис. 2.4). Збіг чи розбіжність запитом визначається самим останнім бітом результуючого вектора R.

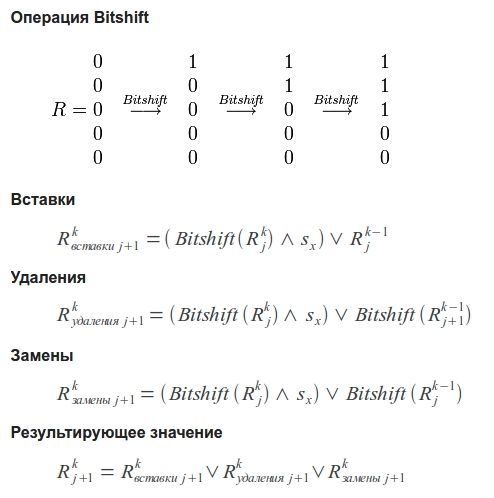


Рисунок 2.4 – Етапи та операції методу

де k - кількість помилок,

j - індекс символу,

sx - маска символу (в масці поодинокі біти розташовуються на позиціях, що відповідають позиціям даного символу в запиті).

Висока швидкість роботи цього алгоритму забезпечується за рахунок битового паралелізму обчислень - за одну операцію можливо провести обчислення над 32 і більше битами одночасно (в залежності від розрядності шини).

При цьому тривіальна реалізація підтримує пошук слів довжиною не більше 32. Це обмеження обумовлюється шириною стандартного типу int (на 32-бітних архітектур). Можна використовувати і типи великих розмірностей, але це може в деякій мірі уповільнити роботу алгоритму.

Не дивлячись на те, що асимптотическое час роботи цього алгоритму O(kn) збігається з таким у лінійного методу, він значно швидше при довгих запитах і кількості помилок k більше 2.

ВИСНОВКИ

ПЕРЕЛІК ПОСИЛАНЬ

1. Phillip, J. Durst and Wendell Gray / J. Phillip // Levels of Autonomy and Autonomous System Performance Assessment for Intelligent Unmanned Systems. – 2014.

2. Keith Thompson. Sales Automation Done Right / Keith Thompson: SalesWays Press – 2005. –296 с.

3. Производственно-сбытовая деятельность [Електронний ресурс] – Ре-жим доступу: www / URL: http://www.ai08.org/index.php/term/ – 02.09.2016 г. – Загл. з екрану.

4. Баркан Д.И. Управление продажами. / Д.И. Баркан – Санкт-Петербург: Высшая школа менеджмента, 2008. – 318 с.

5. Ефимова С.А. Управление сбытом или как увеличить объем продаж / С.А. Ефимова. – Москва: Издательство «Альфа-Пресс», 2007. – 208 с.

6. Аверин Г.В. Системодинамика / Г.В. Аверин. – Донецк: Донбасс, 2014. – 403 с.

7. Manning C. Introduction to Information Retrieval. / Manning C., Raghavan P., Schütze H. – Cambridge University Press, 2008. –198 с

8. Балабанов И. Т. Торговля через виртуальный магазин [Текст] / И. Т. Балабанов // Электронная коммерция, 2004. – С. 195-197.

9. Ашманов И. С. Продвижение сайта в поисковых системах. / И. С. Ашманов, А. А. Иванов – Москва: Вильямс, 2007. – 304 с.

10. Байков В.Д. Интернет. Поиск информации. Продвижение сайтов. / В.Д. Байков – Санкт-Петербург: БХВ-Петербург, 2000. – 288 с.

11. Ландэ Д. В. Интернетика: Навигация в сложных сетях: модели и алгоритмы. / Д. В. Ландэ, А. А. Снарский, И. В Безсуднов. – Mосква: Либроком (Editorial URSS), 2009. – 264 с

12. F. Crestani and G. Pasi. Soft Information Retrieval: Applications of Fuzzy Set Theory and Neural Networks. in «Neuro-fuzzy Techniques for Intelligent Information Systems», N.Kasabov and Robert Kozma Editors, Physica-Verlag, Springer-Verlag Group , 287—313, 1999.

13.

18. https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D1%81%D1%81%D1%82%D0%BE%D1%8F%D0%BD%D0%B8%D0%B5\_%D0%9B%D0%B5%D0%B2%D0%B5%D0%BD%D1%88%D1%82%D0%B5%D0%B9%D0%BD%D0%B0

19. https://habrahabr.ru/post/114997/

20. https://ru.wikipedia.org/wiki/%D0%94%D0%B2%D0%BE%D0%B8%D1%87%D0%BD%D1%8B%D0%B9\_%D0%B0%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC\_%D0%BF%D0%BE%D0%B8%D1%81%D0%BA%D0%B0\_%D0%BF%D0%BE%D0%B4%D1%81%D1%82%D1%80%D0%BE%D0%BA%D0%B8

21.