

4-24-2019

Analysis of efficiency of software tools optimizing the information reliability of electronic documents in automated control systems

I.I. Jumanov

Samarkand State University, Address: University blv. 15, 140104, Samarkand city, Samarkand region, Uzbekistan, Phone: (+99866)2293558, olimjondi@mail.ru

Kh.B. Karshiev

Samarkand State University, Address: University blv. 15, 140104, Samarkand city, Samarkand region, Uzbekistan, Phone: +998936805400, xusan2005@mail.ru

Follow this and additional works at: <https://uzjournals.edu.uz/ijctcm>

 Part of the [Engineering Commons](#)

Recommended Citation

Jumanov, I.I. and Karshiev, Kh.B. (2019) "Analysis of efficiency of software tools optimizing the information reliability of electronic documents in automated control systems," *Chemical Technology, Control and Management*: Vol. 2019 : Iss. 2 , Article 5.

Available at: <https://uzjournals.edu.uz/ijctcm/vol2019/iss2/5>

This Article is brought to you for free and open access by 2030 Uzbekistan Research Online. It has been accepted for inclusion in Chemical Technology, Control and Management by an authorized editor of 2030 Uzbekistan Research Online. For more information, please contact sh.erkinov@edu.uz.

Analysis of efficiency of software tools optimizing the information reliability of electronic documents in automated control systems

Cover Page Footnote

Tashkent State Technical University, SSC «UZSTROYMATERIALY», SSC «UZKIMYOSANOAT», JV «SOVPLASTITAL», Agency on Intellectual Property of the Republic of Uzbekistan

Erratum

?????



ISSN 1815-4840

CHEMICAL TECHNOLOGY. CONTROL AND MANAGEMENT

2019, №2 (86) pp.57-66

International scientific and technical journal
journal homepage: <https://uzjournals.edu.uz/ijctcm/>



Since 2005

УДК 519.681.5

ANALYSIS OF EFFICIENCY OF SOFTWARE TOOLS OPTIMIZING THE INFORMATION RELIABILITY OF ELECTRONIC DOCUMENTS IN AUTOMATED CONTROL SYSTEMS

I.I.Jumanov¹, Kh.B.Karshiev²

^{1,2}Samarkand State University

Address: University blv. 15, 140104, Samarkand city, Samarkand region, Uzbekistan

E-mail: ¹olimjondi@mail.ru, Phone: (+99866)2293558; E-mail: ²xusan2005@mail.ru, Phone: +998936805400

Abstract: The problem has been formulated and conceptual approaches have been proposed to increase the information reliability based on the mechanisms for extracting and using natural, structural-technological, semantic information redundancy of electronic documents. Methods and algorithms have been developed to increase the information reliability based on the use of standard texts of a document, collections of terms, keywords, concepts, a test set of standards, a template, a database (DB) and a knowledge base (KB). The analysis of effectiveness of relevance evaluation algorithms, the errors of the relevance of documents have been provided. A generalized algorithm for improving the reliability of information was tested with search engines for a complete enumeration of variants, heuristic search with annealing, with stochastic modeling based on a truncated Markov chain, and also with tools for adjusting parameters. A methodology has been developed to optimize information reliability and solved a problem, taking into account weights of the effects of each applied control rule. A scheme is proposed, the hardware of which is based on the use of traditional information technology, a local computer network and a database, the capabilities of which are compared with the parameters of the developed information processing algorithms.

Keywords: information accuracy, relevance of document, accuracy of information processing, interconnectedness of elements, concept, a set of standards, a knowledge base, tuning, a generalized algorithm.

Аннотация: Электрон ҳужжат информациясидаги табиий, таркибий-технологик, семантик ортиқчаликларни ажратилиши ва қўллаш механизмлари асосида ахборот ишончилигини ошириши муаммоси талқин этилган ҳамда концептуал ёндашувлари тақлиф қилинган. Ҳужжатдаги типик матнлар, иборалар тўплами, калитли сўзлар, концептлар, назорат эталон – жамламаси, шаблон, маълумотлар базаси (МБ) ва билимлар базасидан фойдаланишига асосланган ахборот ишончилигини оширувчи усул ва алгоритмлар ишлаб чиқилган. Ҳужжат релевантлиги, релевантлик хатолигини баҳоловчи алгоритмлар самарадорлиги таҳлил қилинган. Барча вариантлар бўйича тўла қидириши, юмшоқ эвристик асосда қидириши, Марков кесик занжири бўйича стахостик моделлаштириши асосида қидириши механизмлари, ҳамда параметрларни таҳрир қилувчи воситага эга ахборот ишончилигини оширувчи умумий алгоритм жорий этилган ва тестлаштирилган. Информация ишончилигини оптималлаштириши услубияти тақлиф этилган ва масала, ҳар бир назорат қилувчи қайданинг миқдорий таъсирини инобатга олиб, ечилган. Аппарат қисми, локал ҳисоблаш тармоғи ва МБдан фойдаланувчи, анъанавий ҳамда ишлаб чиқилган маълумотларга ишлов берувчи алгоритмлар имкониятларини солиштирувчи схема тақлиф этилган.

Таянч сўзлар: ахборот ишончилиги, ҳужжат релевантлиги, ахборотга ишлов бериши аниқлиги, элементлар ўзаро боғланиши, концепт, жамлама – эталон, билимлар базаси, сошлаш, умумий алгоритм.

Аннотация: Сформулирована проблема и предложены концептуальные подходы повышения достоверности информации на основе механизмов извлечения и использования естественной, структурно-технологической, семантической информационной избыточности электронных документов. Разработаны методы и алгоритмы повышения достоверности информации, основанных на использовании типовых текстов документа, сборников терминов, ключевых слов, концептов, проверочного набор – эталона, шаблона, базы данных (БД) и базы знаний (БЗ). Проведен анализ эффективности алгоритмов оценки релевантности, ошибок релевантности документов. Протестирован обобщенный алгоритм повышения достоверности информации с механизмами поиска полным перебором вариантов, эвристического поиска с отжигом, со стохастическим моделированием на основе усеченной цепи Маркова, а также инструментами корректировки параметров. Разработана методика оптимизации достоверности информации и решена задача с учетом весов влияния каждого, применяемого правила контроля. Предложена схема, аппаратная часть которой основывается на

применении традиционной информационной технологии, локальной вычислительной сети и БД, возможности которой сравнены с параметрами разработанных алгоритмов обработки информации.

Ключевые слова: достоверность информации, релевантность документа, точность обработки информации, взаимосвязанность элементов, концепт, набор – эталон, база знаний, настройка, обобщенный алгоритм

Актуальность темы

Документы, подлежащие к передаче в автоматизированных системах управления производственно – технологическими комплексами характеризуются высокой вариативностью содержания, в них фиксируется информация, соответствующая различным событиям, которые также связаны с выполнением задач для повышения сохранности, целостности, достоверности, релевантности, точности обработки документов в ситуациях ограниченности априорных сведений, наличия неопределенности и воздействия внешних условий [1-3].

Информация электронных документов (ЭД) искажается при сканировании и распознавании, подготовке к передаче по вине человека - оператора, под влиянием помех в каналах связи, из-за неисправностей, сбоев электронных средств и оборудования, которые становятся источниками ошибок различных типов [4].

Наиболее распространены методы и средства, использующие ошибки обнаруживающих и корректирующих кодов, которые повышают достоверность информации лишь только, в процессе передачи по каналам связи. Наряду с этим, эффективно применяются методы, ориентированные на реализацию протоколов передачи данных на Прикладных и Представительских уровнях многоуровневой сетевой модели OSI [5].

Анализ современной литературы показывает, что в настоящее время исследования по разработке информационных технологий обработки информации на основе методов повышения достоверности в системах электронного документооборота (СЭД) ведутся достаточно интенсивно. Однако, проблема исследования и разработки методов, обеспечивающих до требуемого уровня полноты, релевантности, достоверности и точности обработки ЭД на основе программных инструментов в приемлемых пределах временных и материальных затрат не полностью решена и требует дальнейшего совершенствования существующих и создания новых эффективных методов [6-8].

В настоящей работе предложены концептуальные подходы, принципы и методики расчета количественных и качественных показателей СЭД для повышения достоверности информации, которые основаны на методы, алгоритмы и механизмы извлечения и использования естественной, структурно - технологической, семантической избыточности [4].

Концептуальные принципы и методы повышения достоверности информации на основе информационной избыточности

В табл. 1, 2 даны методики расчета критериальных параметров СЭД, используемых при разработке и тестировании новых алгоритмов повышения достоверности информации [1-3, 6,7].

Предложены и реализованы методы и алгоритмы повышения достоверности информации, основанных на использовании типовых текстов документа, сборников терминов, ключевых слов, концептов, проверочного набор – эталона, шаблона, базы данных (БД) и базы знаний (БЗ) с помощью которых производятся корректировка искаженной информации.

Разработаны методы, использующие изменения фрагментов текстов, составных частей, других особенностей и свойств документов. Причем, алгоритмы, основанные на применении специализированных карт, шаблонов, механизмов по аналогии текстовых редакторов значительно сокращают время контроля целостности, сохранности и достоверности информации [9]. Для расширения возможностей СЭД реализованы механизмы, использующие следующих инструментов повышения достоверности информации на принципах применения: морфологического, синтаксического, графематического анализ [10]; n-граммных структурных моделей естественного языка [11]; логических критериев, семантических связей между элементами и отношений концептов документа [12]; искусственной, естественной, семантической, структурно-технологической информационной избыточности документа [4,9].

Таблица 1
Методики для определения параметров алгоритмов обработки информации СЭД

№	Параметры	Математические выражения	Описание обозначений
1	2	3	4
1	Достоверность Точность Полнота	$\text{Accuracy} = (a + d) / (a + b + c + d)$ $p = \frac{a}{a + c}$ $r = \frac{a}{a + b}$	<p>a - количество, релевантных документов; b - количество, не релевантных документов; c - количество, не найденных релевантных документов; d - количество, найденных и нерелевантных документов.</p>
2	Релевантность документа	$V_{pref} = \frac{1}{R} \sum_{r=1}^N \left(1 - \frac{\text{NonRelBefore}(r)}{R} \right)$	<p>R - число релевантных документов; N - количество документов в коллекции. n - количество запросов; r_i - число, задействованных документов; q_i - число, обработанных документов; error_i - погрешность релевантности</p>
3	Средняя релевантность	$\text{Prec}(r_i) = \frac{1}{N} \sum_{i=1}^N p(r_i, q_i).$	
4	Устойчивость	$\text{MetrikaEr} = \sum_{i=1}^N \text{RelUser}_i - \text{RelPS}_i $	
5	Ошибка релевантности	$E = \sum_{i=1}^n \text{error}_i$	

Таблица 2
Методики разработки механизмов для регулирования параметров алгоритма обработки информации СЭД

№	Параметры	Математические выражения	Описание обозначений
1	Кoeffициент расхождения	$K_{расх} = \frac{\sum_{i=1}^{N_1} D_i(\omega_i) - \sum_{i=1}^{N_2} D_i(\omega_j)}{\sum_{i=1}^N D_i(\omega_i)} = \frac{K_1 - K_2}{K}$	<p>$K_1 = \sum_{i=1}^{N_1} D_i(\omega_i)$ - числа документов с весовым элементом ω_i, которые считаются достоверными;</p> <p>$K_2 = \sum_{i=1}^{N_2} D_i(\omega_j)$ - числа документов с весовым элементом ω_j, которые считаются недостоверными т. е. ($\omega_i \neq \omega_j$);</p> <p>$K = \sum_{i=1}^N D_i(\omega_i)$ - общее число документов, представленных к проверке достоверности элементов;</p> <p>N - общее количество документов.</p>
	Оптимизация алгоритма	<p>с учетом внешних, собственных и внутренних факторов:</p> $Rel(D_i, Q_j) \longrightarrow (SF, DF_Out, DF_inner, GF, OF) \rightarrow \max$ <p>с учетом статистичности и динамичности факторов:</p> $SF(Document) = \sum_{i=1}^{count_sf} k_{sf_i} \cdot sf_i;$ $DF_Out(Document, Query) = \sum_{i=1}^{in_count_dfo} k_{dfo_i} \cdot DF_Out_{ij};$ $DF_inner(Document, Query) = \sum_{i=1}^{count_dji} k_{dji_i} \cdot DF_inner_i;$ $OF(Document, Query) = \sum_{i=1}^{count_of} k_{of_i} * OF_i$	<p>($SF, DF_Out, DF_inner, GF, OF$) - факторы, влияющие на релевантность документа;</p> <p>$count_sf$ - количество статистических факторов;</p> <p>k_{sf_i} - весовой коэффициент внутреннего фактора $[0, k_{max}]$;</p> <p>k_{max} - задаваемое значение весового коэффициента;</p> <p>k_{dfo} - весовой коэффициент внешнего фактора;</p> <p>k_{dji} - весовой коэффициент собственного фактора;</p> <p>sf_i - статистичность фактора.</p> <p>$count_dfo$ - количество внешних факторов;</p> <p>DF_Out_{ij} - значение весового коэффициента внешнего фактора;</p> <p>$count_dji$ - количество внутренних факторов;</p> <p>DF_Inner_i - значение весового коэффициента внутреннего фактора;</p> <p>$count_of$ - количество собственных факторов;</p> <p>OF_i - значение весового коэффициента собственного фактора.</p>

Методика анализа эффективности алгоритмов повышения достоверности информации

Для тестирования разработанных алгоритмов и программных инструментов созданы две базы документов. Первая база документов, используется при обучении алгоритмов, а вторая при оценке критериев релевантности и ошибки релевантности документа. Для чего, сформирован псевдослучайный набор, состоящий из 1000 документов, а также набор, состоящей из 50 различных запросов пользователя системы.

Результаты обработки документов по запросам пользователя определяются временем обращения к БД, БЗ, поиска документов, контроля достоверности и обработки информации по различным задачам. Получены оценки показателя релевантности, ошибок релевантности, как для отдельного, так и для всей коллекции документов, представленных в виде набора с различным размером.

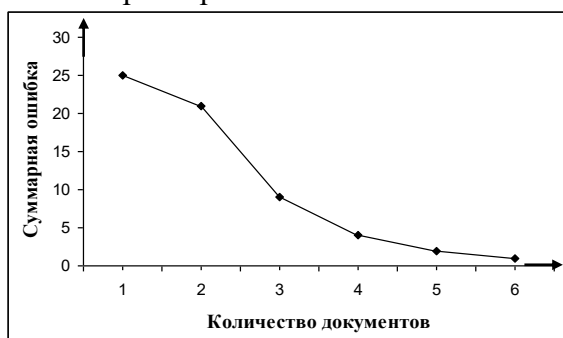


Рис. 1. Ошибка релевантности документа.

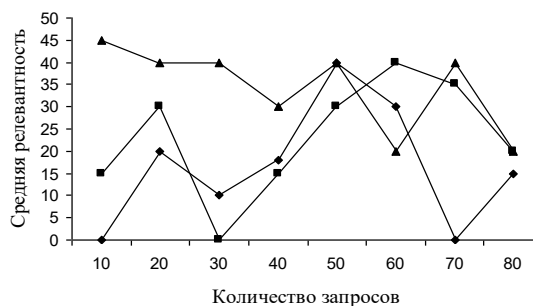


Рис. 2. Релевантность документа при различных механизмах поиска.

На рис.1. показан график суммарной ошибки релевантности документа в зависимости от размера набора документов. Определено, что когда время на расчет показателя релевантности одного документа равно в среднем 2-3 мин., тогда для обработки заданного псевдослучайного набора из 1000 документов потребуется примерно 90-125 час. Наименьшее значение ошибки релевантности достигается при наборе из 100 документов. Достоверность расчета релевантности документа оценивается с вероятностью $\mathfrak{R} = 0,95$.

Протестирован алгоритм повышения достоверности информации, который синтезирует механизмы поиска с полным перебором вариантов, эвристического с отжигом, со стохастическим моделированием на основе усеченной цепи Маркова. Синтезированы также стандартные инструменты, включающие встроенные средства корректировки коэффициентов, подбираемых при адаптации параметров СЭД.

Определено, что эффективность обобщенного алгоритма достигает максимальной точки при числе 20-30 запросов, а в наборах с большим числом запросов наблюдается резкое уменьшение значения показателя релевантности документа. На рис. 2, показана графическое сопоставление средней релевантности документа по обобщенным алгоритмом, где \blacklozenge линией показан результат синтеза механизма поиска с полным перебором, \blacksquare линией показан результат механизма поиска с отжигом, \blacktriangle линией показан результат механизма поиска со стохастическим моделированием.

На рис. 3 показан график ошибок релевантности документа по обобщенному алгоритму, отражаемая зависимостью от последовательности числа запросов, где \blacksquare обозначен график на основе механизма поиска с полным перебором, \blacktriangle график на основе механизма поиска с отжигом, \times - график поиска на основе стохастического моделирования.

Эффективность обобщенного алгоритма исследована по критерию F- меры, которая объединяет другие метрики расстояний при оценке показателей полноты и точности обработки документов. Установлено, что значение показателя релевантности повышается в среднем на 40% при механизме поиска с полным перебором, на 55% при механизме поиска с отжигом, на 70% при механизме поиска со стохастическим моделированием.

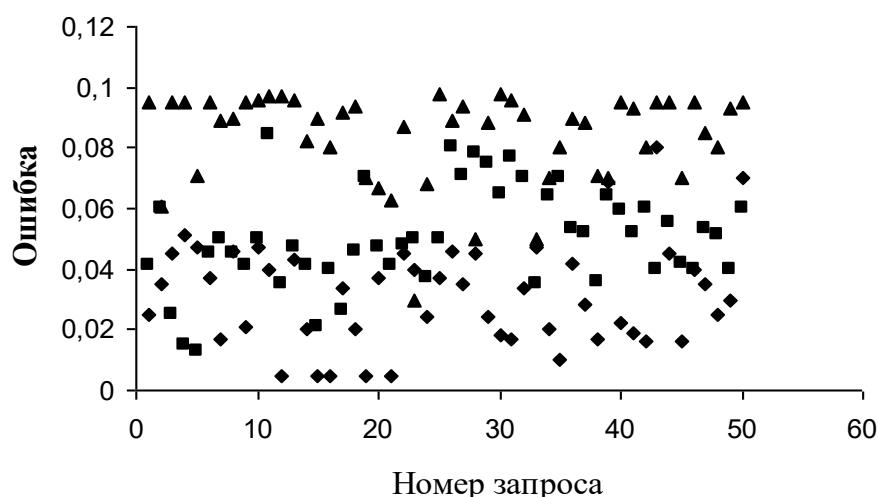


Рис. 3. Ошибки релевантности документа в зависимости от числа запросов.

В табл.3 даны экспериментальные значения для параметров тестируемого обобщенного алгоритма повышения достоверности информации. В механизмах расчетов использованы следующие переменные: tf – частота, встречаемости ключевого концепта в документе; idf – обратная частота встречаемости $idf = \lg \frac{N}{Df}$; w_{tq} – весовое значение элемента ключевого концепта в документе $w_{t,q} = tf * idf$; w_f – весовое значение ключевого концепта в документе; $w_{t,d}$ – нормированное весовое значение ключевого концепта документа; Df – трудоёмкость обработки информации ЭД.

Эффективность обобщенного алгоритма исследована по критерию трудоёмкости обработки информации ЭД.

Таблица 3

Экспериментальные значения параметров обобщенного алгоритма

Ключевые концепты	Запрос				Документ			Скалярное произведение
	tf	Df	idf	$W_{t,q}$	tf	w_f	$w_{t,d}$	
секретариат	0	5000	2,3	0	1	1	0,41	0
учебный отдел	1	50000	1,3	1,3	0	0	0	0
информатика	1	10000	2,0	2,0	1	1	0,41	0,82
прикладная математика	1	1000	3,0	3,0	2	2	0,82	2,46

Определено, что на порядки уменьшается значение коэффициента трудоёмкости обработки информации по обобщенному алгоритму с механизмом поиска ключевых концептов документа и достигается: $O(N^4)$ оп/сек при полном переборе; $O(N^3)$ оп/сек при поиске с отжигом; $O(N^2)$ оп/сек при поиске со стохастическим моделированием; $O(N)$ оп/сек при поиске по семантической сетевой модели.

На рис. 5 а) показан график средней точности обработки информации по обобщенному алгоритму с механизмами поиска с отжигом и со стохастическим моделированием в зависимости от числа запросов, а на рис. 5 б) ошибка средней точности обработки информации.

Методика оптимизации показателя достоверности информации

Элемент концепта ЭД x представляется в диапазоне значений от 0 до k . Динамика его изменения описывается дифференциальным уравнением вида

$$\frac{dx}{dt} = uk - v.$$

где k - коэффициент динамики изменения элемента концепта ЭД; u - параметр внутреннего воздействия на процесс обработки информации; v - параметр внешнего воздействия на процесс обработки информации.

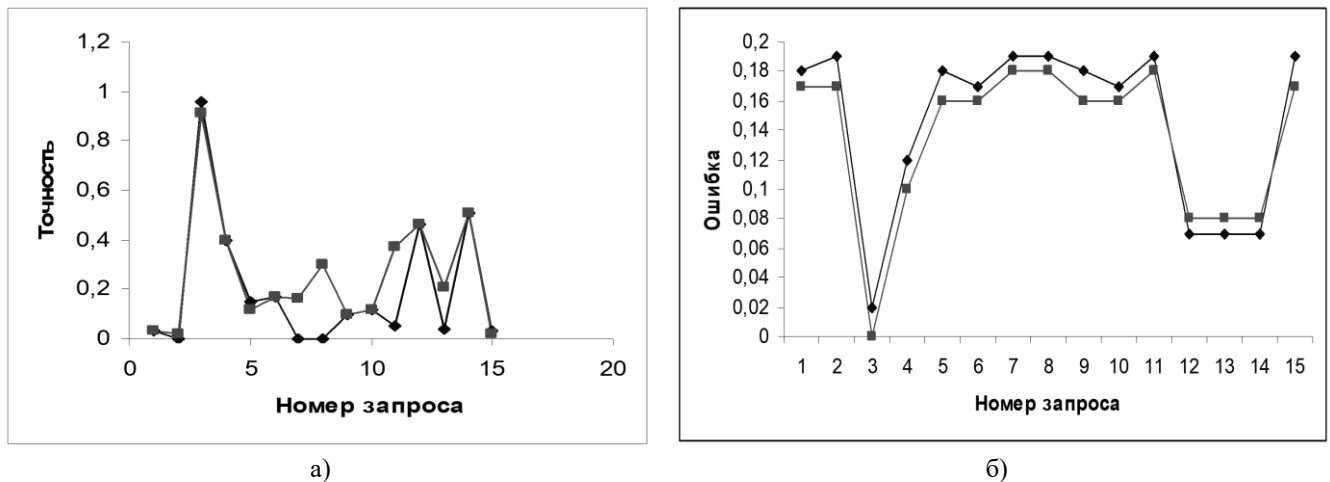


Рис. 5. Эффективность обработки информации в зависимости от номера запросов: а) точность обработки информации; б) ошибка в точности обработки информации.

Функциональные качества обработки информации представляются в виде

$$Q = \int_0^T (x(t) - g)^2 dt \rightarrow \min ,$$

где g - порог, выбираемый в интервале значений $[0, \dots, 1]$.

Предлагаемый принцип для оптимизации достоверности информации направлен на применение механизма регулирования выходного результата алгоритма с целью поддержания его значения на требуемом уровне. Минимизируется среднеквадратическое отклонение значения элемента концепта $x(t_i)$ в t_i момент времени по результатам сравнения с порогом, значение которого заложено в проверочном наборе - эталоне.

При этом, правило алгоритма предполагает сравнение значения элемента концепта ЭД $x(t_i)$ со значениями характеристик набора - эталона с целью выявления разности, которая формируется в виде бинарной последовательности 0 и 1, где «0» - соответствует недостоверному, а «1» - достоверному элементу.

Отметим, что методика оптимизации задачи основывается на задании условия Коши $x(0) = 0$. В таком случае, известный принцип максимума Понтрягина неприменим так, как при достижении $x(t)$ порогового уровня g могут существовать бесконечное количество переключений. В связи с этим, предложены для применения эвристические подходы к оптимизации.

Вариант 1. Контроль достоверности осуществляется на основе правил:

- если $x < g$ то $u = 1$;
- если $x = g$ то $u = v/k$;
- если $x > g$ то $u = 0$.

Элемент x представляется в реальном режиме времени с определенным интервалом значений. Если в принятом интервале значение контролируемого элемента превышает уровень порога g , то считается, что значение расчетного элемента недостоверно.

Вариант 2. Контроль достоверности осуществляется по адаптивным правилам:

- если x "меньше g " то $u = 1$;

- если $x = "g"$ то $u = v/k$;
- если $x = "больше g"$ то $u = 0$.

Вариант 3. Контроль достоверности производится по «мягким» правилам:

- если $x = "меньше g"$ то $u = "около 1"$;
- если $x = "g"$ то $u = "около v/k"$;
- если $x = "больше g"$ то $u = "около 0"$.

Эффективность алгоритма, в первую очередь, зависит от степени принадлежности элемента концепта подмножествам достоверных D_i и недостоверных T_i значений. Оптимизация основывается на механизм регулирования значений параметров u , v и k , которые способствуют положительному изменению значения показателя эффективности E_0 .

Представляет также интерес решение задачи оптимизации с учетом весовых вкладов каждого, применяемого правило в структуре алгоритма повышения достоверности информации, задаваемых в виде зависимости

$$\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_N\},$$

где $\lambda_1, \lambda_2, \dots, \lambda_N$ – веса вкладов правил для повышения достоверности информации.

Исследование показателя E_0 проводится по соотношению

$$E_0 = \lambda D_i + (1 - \lambda) T_i,$$

λ – интенсивность запроса, оцениваемая по вероятностной функции принадлежности элемента концепта подмножествам D_i и T_i .

Квадратичное расстояние близости элементов из подмножеств достоверных значений i и недостоверных значений j с учетом весов вклада правил алгоритма задается, как

$$d_k^2(\omega_i, \omega_j) = \lambda_1 d_1^2(\omega_i, \omega_j) + \lambda_2 d_2^2(\omega_i, \omega_j) + \dots + \lambda_{10} d_{10}^2(\omega_i, \omega_j),$$

где ω_i, ω_j – соответственно, весовые значения элементов i и j .

Значимость меры близости оценивается по границам нормального закона распределения, критериев Колмогорова-Смирнова, Романовского, χ^2 – Пирсона и др.

Эффективность варианта обработки информации без правила повышения достоверности элемента в концепте оценивается в диапазоне $B = \alpha_g - \alpha_n$ с нижней α_n и верхней α_g границами по значениям следующего коэффициента

$$k = \frac{1}{n} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left(\frac{|d(\omega_i, \omega_j)|}{B} \right),$$

где $B = m^n$ – диапазон изменения значений j -го элемента концепта ЭД; m – основание кода, в данном случае, десятичного кода; n – разрядность десятичного кода.

Эффективность алгоритма повышения достоверности информации оценивается по коэффициенту изменения элемента k_d или же размера концепта в виде

$$k_d = \lambda_0 + \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \lambda_j d_j^2(\omega_i, \omega_j).$$

Для реализации алгоритмов предложена схема, аппаратная часть которой основывается на применение традиционных средств информационной технологии, локальной вычислительной сети и БД. Схема учитывает возможность алгоритма на основе контроля

робастности кадров оцифрованных изображений документа на основе шаблона, набора – эталона фрактальных характеристик, которые заранее заложены в БЗ.

Для тестирования оптимизированной схемы повышения достоверности информации ЭД сформирован набор, включающий 200 кадров изображений 12 оригиналов документов. Средняя вероятность ошибок информации, происходящая на этапах ввода, передачи, хранения, обработки документа задана $P = 10^{-3} \div 10^{-4}$. Коэффициент повышения достоверности информации задан $k_d = 0,98$.

В табл. 4 сравнены результаты существующей технологии и предложенной, включающие метод повышения достоверности информации на основе набора – эталона.

Таблица 4

Эффективность существующей и предложенной технологии повышения достоверности информации

Документы	Существующая технология			Предложенная технология	
	Время обработки информации (в сек)	Время контроля достоверности ЭД (в сек)	Время представления ЭД пользователю (в сек)	Время контроля достоверности ЭД (в сек)	Время представления ЭД пользователю (в сек)
1	20,5	104,5	48	19,12	19,2
2	22,5	110	50	20,37	45,37
3	18,5	111	46	20,14	45,14
4	20,5	103	48	19,83	44,83
5	14,5	71,5	42	12,94	37,94
6	15,5	73,5	43	12,53	37,53
7	16	69	43,5	12,78	37,78
8	16,5	68,5	44	12,48	37,48
9	16	72,5	43,5	13,78	38,78
10	16	73	43,5	13,91	38,91
11	15,5	72,5	43	13,78	38,78
12	16	73	43,5	13,85	38,85
Среднее время:			44,83		40,46

Заключение

В результате тестирования алгоритмов СЭД получены следующие усредненные значения параметров: количество документов, стоящих в очереди на индексацию - 4; время индексации документа - 3 сек; число обработанных документов по запросам - 20; время ответа на запросы пользователя - 2 сек. ; общее количество документов, подлежащих индексации - 10 000; ёмкость документа - 2 500 элементов и 15 концептов.

Время на поиск, обнаружения и исправления ошибок документа уменьшается с 4.39 мин до 3.3 мин. Затрачивается на 30% меньше времени, чем на поиск и обработки ЭД с визуальным контролем достоверности информации.

53% пользователей получают результат, после первого же запроса; 47% пользователей дают в среднем 4 запроса. Средняя погрешность выполнения комплекса составляет 7,25%, а число повторных запросов с целью повышения достоверности ЭД уменьшается на 50% .

Время нахождения искаженной информации в ЭД сокращается на 65%; время на передачу, поиск, распознавание, классификацию и контроль достоверности информации в целом по коллекции ЭД сокращается на 40%; загрузка на сервер СЭД снижается на 6%.

References:

1. Keldysh N. V. Metodicheskie osnovy avtomatizirovannogo resheniya zadach vedomstvennogo elektronnoho dokumentooborota. // Nauchno-metod. sbornik № 56. / VA MO. - M., 2009. - S.110-117 Inv. № 58592.
2. Gudov, A. M. Ob odnoj modeli optimizacii dokumentopotokov, realizuemoj pri sozdanii sistemy elektronnoho dokumentooborota vychislitel'nye tekhnologii. - 2006. - T. GG. spec. vyp. - S. 53 - 65.
3. Bessonov, S. V. Optimizaciya elektronnoho dokumentooborota v korporativnyh sistemah: dis. kand. ekon. nauk. M., 2000 g. 187 s.
4. ZHumanov I.I. Konceptual'nye principy i metody kontrolya dostovernosti informacii v strukture paketov peredachi dannyh na osnove statisticheskoj izbytochnosti // «Ilmij tadqiqotlar ahborotnomasi» ilmij-nazarij, uslubij zhurnal. – Samarqand: SamDU, 2013. - №1 (77) – 39-49 b.
5. ZHumanov I.I., Karshiev H.B. Metody obespecheniya dostovernosti elektronnyh dokumentov na osnove strukturnoj izbytochnosti i leksikologicheskogo sinteza // «Nauka i mir», Mezhdunarodnyj nauchnyj zhurnal, Izd-vo «Nauchnoe obozrenie», Volgograd. – №3(55), Tom 1, 2018.– s. 49-51.
6. Norenkov I.P., Uvarov M.YU. Baza i generator obrazovatel'nyh resursov // Informacionnye tekhnologii, 2005, № 9, s. 60-65.
7. Gavrilova T. A., Horoshevskij V. F. Bazy znaniy intellektual'nyh sistem. SPb: Piter, 2000. 384 s.
8. Lukashevich N. V. Tezaurusy v zadachah informacionnogo poiska. M.: Izd-vo Moskovskogo universiteta, 2011. 512 s.
9. ZHumanov I.I., Karshiev H.B. Expanding the possibilities of instruments to improve the information reliability of electronic documents of industrial management SYSTEMS // Tenth World Conference “Intelligent Systems for Industrial Automation”, WCIS-2018, 25-26 October 2018, Tashkent, Uzbekistan, ISBN: 933609-37-2-2018, , b-Quadrat Verlag-86916 Kaufering, –312-316 p.
10. ZHumanov I.I., Ahatov A.R. Ocenki dostovernosti peredachi izobrazhenij elementov teksta v telekommunikacionnyh setyah// NTZH «Himicheskaya tekhnologiya. Kontrol' i upravlenie» -TGTU, Tashkent, 2010 - № 2.- s. 23-30.
11. ZHumanov I.I., Ahatov A.R. Optimizaciya kontrolya peredachi i obrabotki informacii na baze tekhnologii parallel'nyh vychislenij SUDA// NTZH «Himicheskaya tekhnologiya. Kontrol' i upravlenie» - TGTU, Tashkent, 2009- № 5, s. 33-39.
12. ZHumanov I.I., Ahatov A.R. Algoritmy kontrolya dostovernosti izobrazhenij elementov teksta v strukture paketov peredachi dannyh// NTZH «Himicheskije tekhnologii. Kontrol' i upravlenie». - Tashkent, 2010. - № 3. - s.39-46.