

# **SoftInform**

## **Search Technology**

## Роль поисковых технологий в информационных системах и системах документооборота

---

Одной из важнейших задач, которую приходится решать компаниям на сегодняшний день является организация быстрого поиска документов в больших объемах данных

Организация доступа к данным напрямую зависит от технологий и программ, обеспечивающих скорость и качество обработки информации.



Наш опыт работы с корпоративными потребителями показал, что на крупном предприятии существует ряд проблем, напрямую связанных с поиском информации.

## Типовые проблемы корпоративных пользователей



**Сокращение времени поисковой сессии**



**Размытость информационного наполнения**



**Консолидация информации из различных источников**



**Первичный аудит информационных потоков**

## Сокращение времени поисковой сессии

### Фразовый поиск. Стандартная поисковая сессия

- Подбор ключевых слов
- Просмотр результатов
- Повторный подбор ключевой фразы и поиск
- Etc...



Например ищем информацию по покупке компаний. По ключевой фразе «покупка компаний» получаем список документов. После просмотра результатов поисковая сессия продолжается (так как не всегда первые результаты удовлетворительны). Снова подбор ключевых слов: «слияние компаний» и «приобретение компаний» и т.д. И в итоге приходится подбирать нужные ключевые фразы для поиска и перерабатывать большое число документов. И что самое важное нет гарантий что мы вспомним все нужные фразы.

**Основные проблемы:** затраты времени на подбор правильных ключевых слов и просмотр ненужных документов.

**Цель:** сокращение времени поисковой сессии

## Проблема размытости информационного наполнения

В базе данных или информационной системе предприятия могут содержаться документы из различных источников, содержащие похожую или идентичную информацию.

Один и тот же текст может быть с разными заголовками, с небольшими изменениями или дополнениями, что вносит определенную путаницу при его использовании.



Так к примеру один специалист даст комментарии к документу №1, другой специалист - к документу №2. В итоге при поиске случайным образом находится один из документов и видны комментарии только одного специалиста.

## Консолидация информации из разных источников

На этой проблеме необходимо заострить внимание, так как по мере роста информации окружающей нас она становится все актуальней, и крупные предприятия вынуждены затрачивать огромные средства, на совмещение информации из различных систем в единую.

Еще одной проблемой является внедрение на предприятие, что может крайне болезненно сказаться на работе управляющего персонала в течении определенного времени после внедрения.



Технологии поиска информации могут являться консолидирующим элементом для различных информационных систем. Работа не вместо, а в связке с уже установленными системами позволяет структурировать информационные составляющие любого крупного предприятия под управлением одной программы.

## Первичный аудит информационных потоков

Достаточно распространенное явление – содержащиеся в информационной базе предприятия дублирующие друг друга документы из различных источников или внесенные разными людьми.



Так как информация, как правило, накапливалась годами, то чтобы полностью использовать все преимущества новых технологий поиска необходимо вначале избавить систему от ненужных дублей. Как показывает практика после проведения этой работы руководители подразделений «хватаются за голову», видя какой беспорядок был в организации работы с информацией.

Все эти, а также множество других проблем позволяет решать технология поиска компании СофтИнформ.

Основным преимуществом и отличием SoftInform Search Technology от существующих аналогичных технологий и систем поиска является запатентованная компанией СофтИнформ функция поиска документов, похожих по содержанию на текст запроса. Именно эта уникальная возможность технологии СофтИнформ и позволяет наиболее эффективно решать большинство проблем обработки и поиска информации на предприятии



## Сокращение времени поисковой сессии



-Запрос по ключевой фразе

-Просмотр результатов

- сразу как находим первый документ который нам интересен, то просто жмем кнопку "дать похожие по содержанию документы" и сразу же видим релевантный список документов именно на заданную тему.

Таким образом, вместо того чтоб затратить на поиск нужной информации несколько часов (просмотр списков результатов и подбор нужных ключевых фраз), можно, при помощи технологии поиска похожих, управиться за пару минут.

### Избавление от размытости информационного наполнения

Дублирование документов приводит к серьезным временным затратам на работу с ними (просмотр, редактирование и т.д.) что в итоге снижает производительность персонала.



Технологии *поиска похожих по содержанию документов* от СофтИнформ решают эту проблему путем сравнения поступающих в базу предприятия документов с уже содержащимися там данными и выявления дублей.

### Консолидация информации из разных источников

Системы на базе SoftInform Search Technology легко интегрируются в информационную структуру предприятия, подключая различные источники данных, и имеют архитектуру клиент-сервер.



Еще одним преимуществом технологий поиска СофтИнформ является то, что данная система легко и без продолжительной доработки встраивается [надстраивается] в любую информационную систему. Ничего менять (что связано с дополнительными затратами) не надо. Наш поисковик работает совместно с любым установленным на предприятии ПО – будь то CRM системы, СУБД или системы управления знаниями.

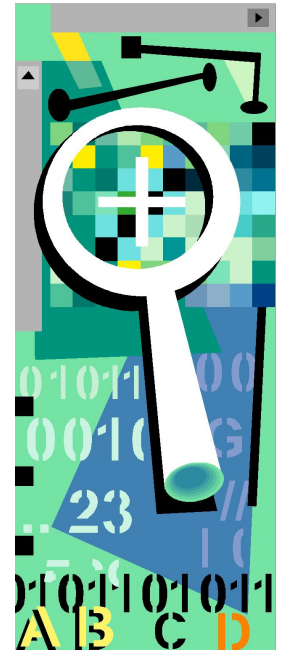
### Первичный аудит информационных потоков

Для определения дублей и ненужных «похожих» файлов можно использовать функцию построения отчета анализа схожести. При этом эта операция занимает в десятки раз меньше времени, чем при обычном сравнении.

Часто дублирующие документы находятся в различных информационных источниках.

Построение отчета о дублирующих друг друга документах – это одна из составляющих общего аудита информационных потоков на предприятии, который могут провести наши специалисты.

После проведения такого аудита мы можем предложить компании наиболее оптимальное решение различного рода проблем, связанных с поиском и структуризацией информации в рамках предприятия.



SoftInform Search Technology – технология поиска документов, похожих по содержанию в текстовых файлах практически любых форматов, базах данных и информационных системах

## Возможности SoftInform Search Technology



- Уникальный поиск похожих документов
- Скорость индексации от 15 до 30 Гб/час
- Малый размер индекса (20-25% от объема индексируемой информации)
- Поддержка практически всех распространенных форматов текстовых файлов
- Концепция различных источников данных
- Масштабируемость
- Консолидация информации на предприятии

SoftInform Search Technology является языковисимой. Все, что связано с каким-либо языком монтируется дополнительными модулями.

## Наши партнёры

---

Технологии компании «СофтИнформ» получили признание от многих именитых и уважаемых компаний, которые уже не первый год успешно функционируют на IT-рынке.

Среди наших партнёров находятся такие гранды как **IBM** и **ABBYY**.

Результатом подписанного с **IBM** соглашения стало присоединение компании «СофтИнформ» к партнёрской программе IBM PartnerWorld Industry Networks (PWIN). По итогам заключенного соглашения «СофтИнформ» интегрировал своё поисковое решение **SearchInform** с программными продуктами линеек **Lotus** и **WebSphere**.

По итогам соглашения с **ABBYY**, состоялась интеграция системы распознавания отсканированных документов **ABBYY Recognition Server** в контур перевода печатной продукции в электронный вид от компании «СофтИнформ».

На данный момент SoftInform Search Technology уже известна на российском рынке.

Опыт нашей работы подсказывает, что технологии поиска будут востребованы на протяжении многих лет.



### **Best Soft 2005 По версии PC Magazine**

Одним из критериев востребованности полнотекстового поиска как для конечного пользователя, так и [в первую очередь] для предприятий и организаций является повышенный интерес российской IT-прессы к нашей технологии и ПО на ее основе

**Спасибо за внимание**  
**ВОПРОСЫ?**

