
ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА В ПРОЦЕССАХ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА

Д.О. Гринько

Кафедра кибернетики и мехатроники
Российский университет дружбы народов
ул. Орджоникидзе, 3, Москва, Россия, 115419

Статья содержит краткую информацию о ведении электронного документооборота (ЭД) в компаниях с использованием информационных систем на примере системы SAP ERP. Дан анализ основных принципов ведения ЭД и недостатков стандартных решений. Предложен подход по использованию интеллектуального анализа данных для поддержания консистентности системы и контроля за оперативным вводом данных. Приведен пример использования данного подхода.

Ключевые слова: электронный документооборот, САП, информационные системы, интеллектуальный анализ, представление документов в информационной системе, электронный документ.

В настоящее время большинство крупных компаний использует информационные системы (ИС) для автоматизации и оптимизации своей деятельности. Это позволяет минимизировать число ручных операций, снизить риск возникновения ошибок, связанных с человеческим фактором, повысить скорость передачи информации и многое другое. Внедрение ERP-систем обеспечивает компанию возможностью вести весь документооборот в электронном виде. В данной работе будет рассматриваться использование ИС на примере ERP-систем [2].

ИС разрабатываются по модульному принципу, позволяющему провести грань между различными областями деятельности предприятия. При этом каждый модуль остается зависимым от смежных с ним модулей системы. При внедрении ИС определяется перечень автоматизируемых процессов, на основе которых выбираются модули для использования в рамках проекта.

Использование ИС и, в частности, ERP-систем дает множество преимуществ, но сопровождается и рядом проблем. Например, стандартный пакет ERP включает в себя множество моделей процессов, выбор и конфигурация которых осуществляется человеком. Проекты полного цикла (от начала перехода на ERP до полной автоматизации) длятся очень долго, в некоторых случаях более 10 лет. За это время неизбежно меняется штат сотрудников. Это приводит к тому, что процессы конфигурируются разными людьми. Из-за недостаточно полной документации, недостаточной квалификации отдельных сотрудников повышаются риски возникновения ошибок при настройке системы, связанных с человеческим фактором. Это, в свою очередь, ведет к некорректной работе системы в рамках продуктивной эксплуатации и оперативным ошибкам.

Кроме того, практически в каждой компании существует набор специфических процессов, которые не покрываются множеством стандартных моделей. Для них необходимо расширять возможности ERP за счет дополнительных разработок. Разрабатываемые «под проект» дополнительные процессы нуждаются в создании соответствующих механизмов контроля.

Для реализации связей между документами в стандартной поставке ERP-систем обычно используются фиксированные поля в ссылающихся документах, которые задают ключевые поля других объектов системы (других типов документов). Каждому полю-«ссылке» присвоен набор подпрограмм, обеспечивающий корректность его заполнения и влияние его на данные текущего документа.

Особенности обработки ЭД в ИС

Рассмотрим ведение электронного документооборота на примере системы SAP ERP. Каждый модуль объединяет в себе фиксированный набор процессов и оперативных, и основных данных. Для обеспечения процессов каждого модуля и хранения оперативных данных в системе создаются электронные документы (ЭД).

ЭД, в свою очередь, можно условно разделить на два вида:

1) реальные (прообразы бумажных) документов: договоры с контрагентами, бухгалтерские счета-фактуры, банковская выписка, документы заказа товара, таможенные накладные и др.;

2) вспомогательные: документы планирования расходов, вспомогательные счета, лимиты бюджета, технические договора кассового обслуживания.

Для компании определяется набор сквозных процессов — последовательность процессов, которая полностью отражает выполнение типовой операции компании. Для каждого сквозного процесса выделяется цепочка ЭД — последовательность их создания и обработки. Например, на рис. 1 представлена цепочка документов для оплаты аренды: заключение договора с арендодателем, создание заказа на аренду, получение счета и его оплата.



Рис. 1. Цепочка документов для оплаты аренды

Число цепочек документов, соответствующих сквозным процессам компании, меняется со временем и может достигать десятков тысяч. Во время продуктивной эксплуатации ERP возможны различные отклонения процессов в системе от стандартных цепочек ЭД [5].

В каждой компании бывают ситуации, когда оплату необходимо произвести срочно и при этом нет времени на ожидание создания предыдущих ЭД, например, при угрозе отключения электричества на производстве. В этих случаях в виде исключения цепочка ЭД нарушается. Если в дальнейшем подобная «исключительная ситуация» не будет должным образом обработана (воссоздана соответствующая хронология и связь ЭД), то компания рискует получить штрафы и санкции из-за сдачи недостоверной отчетности и т.д.

Оперативные данные ведутся в ЭД и в большинстве случаев представляют собой ссылку на запись основных данных. Основные данные ведутся на специальных карточках, которые не должны изменяться в рамках протекания очередного процесса.

Согласно модульному методу внедрения ERP-систем, число цепочек документов растет по мере расширения автоматизируемых направлений деятельности компании.

Особенности представления ЭД в ИС

Технически все электронные документы одного типа хранятся в одной или нескольких таблицах базы данных. Таблицы, относящиеся к одному типу документов, имеют либо полностью, либо частично совпадающие ключевые поля. Например, в системе SAP ERP основные данные бухгалтерских документов хранятся в двух таблицах базы данных: таблице заголовков и таблице позиций. При этом таблица, хранящая позиции документов имеет те же ключевые поля, что и таблица заголовков, но с одним дополнительным полем — «номер позиции» (рис. 2).

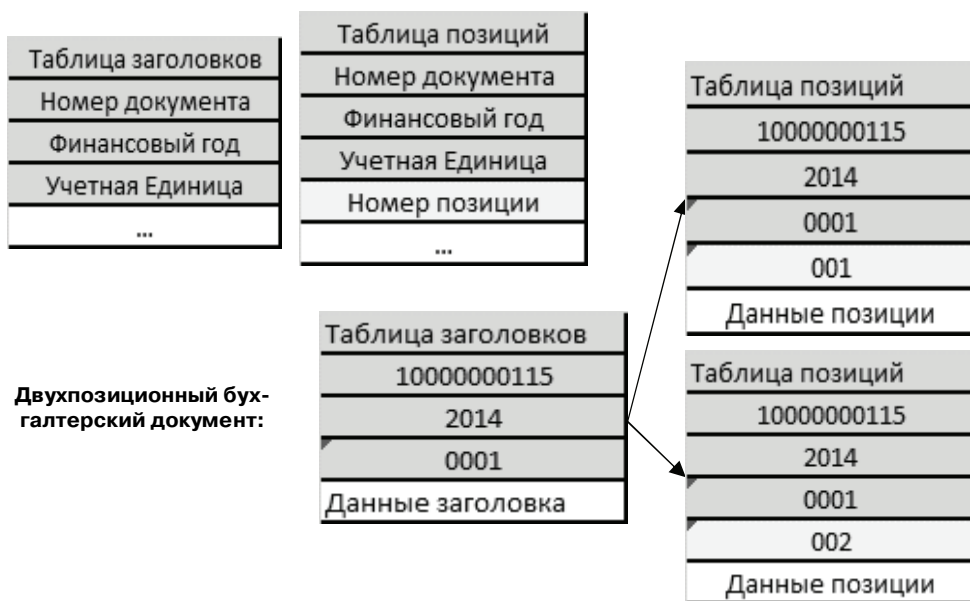


Рис. 2. Стандартный способ хранения основных данных бухгалтерского документа

Для ссылки на другие электронные документы в системе в исходный документ добавляются поля, являющиеся ключевыми для ссылочной таблицы.

Будем далее условно считать, что ЭД разделяются по типам и видам.

ЭД, хранящиеся в одних и тех же таблицах БД, будем объединять в один **тип**. В частности, на примере SAP ERP к разным типам относятся:

Тип 1: Документы бухгалтерского и налогового учета (т. BSEG, BKPF)

Тип 2: Документы планирования платежей (т. KBLP, KBLK)

Тип 3: Документы закупок (т. MSEG, MKPF)

Видами (и подвидами) документов разграничиваются документы одного типа, хранящиеся в общем случае в идентичном наборе таблиц, имеющих общий набор полей, отличающихся только их значениями. На примере SAP ERP к видам документов для типа 1 относятся:

Вид 1: Счета-фактуры (таблицы BSEG, BKPF)

Вид 2: Документы банковской выписки (т. BSEG, BKPF)

Вид 3: Документы амортизации основных средств (т. BSEG, BKPF)

Связанными между собой могут быть как ЭД различных видов и типов, так и одинаковых. В частности, цепочке документов на рис. 3 соответствует набор таблиц БД со ссылками на смежные ЭД из рис. 4. Стандартно в данной цепочке ЭД финансовый документ создается на каждую позицию документа закупки товара/услуги [4].

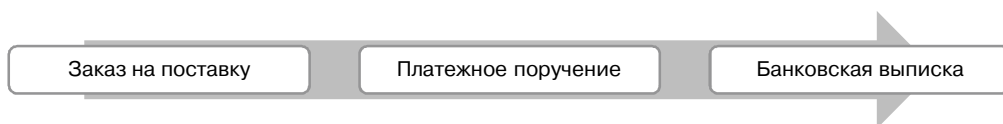


Рис. 3. Цепочка документов

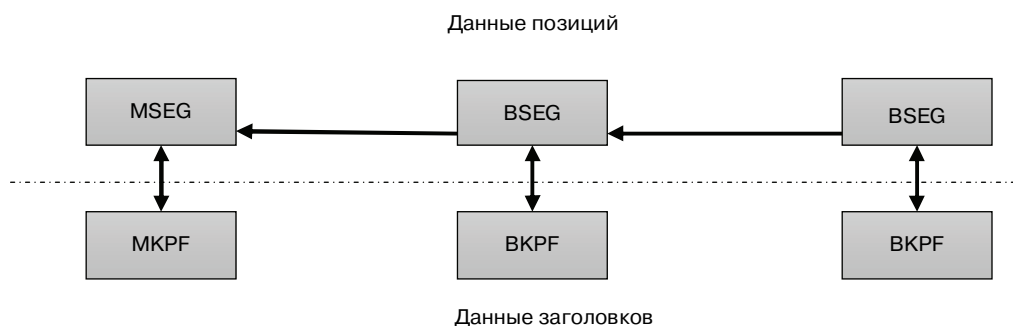


Рис. 4. Стандартное распределение данных и ссылок между ЭД по таблицам БД

Если для компании неважно разбиение документов по позициям (каждому товару/услуге в рамках одного заказа), но важна производительность, то при внедрении стандартная связь между данными позиций может быть изменена на связь между данными заголовков документов (рис. 5).

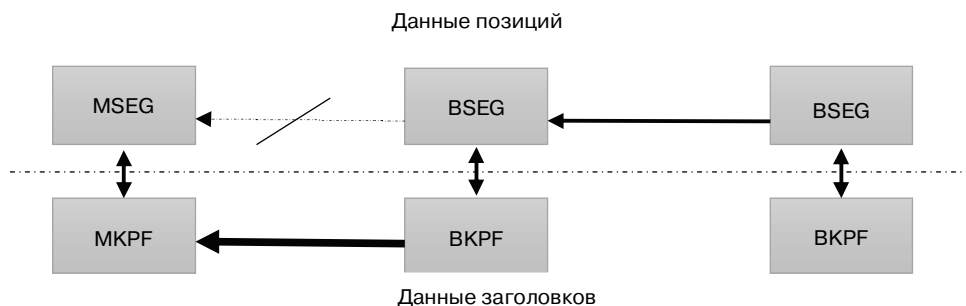


Рис. 5. Нестандартное использование ссылок между ЭД

Недостатки стандартного решения

Внедрение ИС происходит поэтапно. Вначале внедряется несколько наиболее приоритетных модулей и соответствующих им цепочек ЭД, затем следующие по приоритету и так далее. В данной работе ставится задача преодоления следующих недостатков ИС:

1. Ограниченное число связей между документами и их фиксированность
 - а. Большая часть ИС разрабатывается за рубежом. Специфика процессов Российской Федерации практически всегда требует реализации дополнительных типов и видов документов, установки связи между ними
 - б. Ширина структур таблиц базы данных ограничена. Не релевантные для проекта ссылки уменьшают область применимость различных видов документов. Использовать ссылочные поля не по назначению при этом не позволят стандартные функциональные модули системы по проверке заполнения полей
 - с. Множество ссылок в системе реализованы косвенным путем (через третий тип документов). Построение отчетов по косвенным ссылкам работает медленнее, чем по прямым. В случае частого формирования отчетности, использующей косвенные ссылки, может иметь смысл преобразование их в прямые
2. Определение взаимосвязей ЭД в системе человеком (обычно специалистом внедряющей компании)
 - а. Оптимальность реализованных связей в итоге зависит от качества принятия решения специалистом в короткие сроки, без возможности изменения его в дальнейшем (ввиду наполнения базы данных реальными документами, структура которых не может быть изменена в режиме реального времени)
 - б. Качество концепции по реализации связей зависимо от человеческого фактора
3. Невозможность обеспечить четкий алгоритм ведения связей в системе при внедрении из-за большого числа видов документов и операций с ними.

Негативные последствия от перечисленных недостатков могут проявиться в следующем: сдача в государственные органы недостоверной отчетности, необъяснимая с точки зрения ERP-системы нехватка денег на расчетных счетах компа-

нии, повторная оплата кредиторской задолженности, срыв сроков поставки или условий договора и др. (Например, не указанный в документе закупки договор не позволит оценить выполнение обязательств по договору).

Предлагаемое решение

Обозначим за T число различных типов документов, используемых в рамках эксплуатации ИС. Пусть имеется V видов документов, определяемое как

$$V = \sum_{i=1}^K v_i(p_i),$$

где v_i — число видов документов типа k_i , а p_i — перечень полей документов типа k_i .

Для поддержания консистентности системы каждый документ одного вида должен содержать в себе определенное число ссылок на другие документы. Соответственно, для этого необходимо определить множество ссылок к каждому виду документов:

$$L = f(v_i(p_i), t)$$

и поддерживать его постоянство для документов одного вида. При этом число ссылок будет также зависеть и от момента времени (t), так как процессы любой компании подвержены изменениям.

Предлагается подход, реализующий следующие этапы:

1. Интеллектуального анализа таблиц базы данных для нахождения схожих документов, выделения отдельных их видов, определения типичных для данного вида таблиц связей

2. Формирование на основе анализа таблиц базы данных продукционных правил, которые могут быть использованы для контроля ведения ЭД в режиме реального времени

3. Контроль целостности отдельных цепочек документов: отслеживание хронологии создания, заполненности ссылочных полей, наличия и корректности каждого документа из проверяемой цепочки с возможностью аргументированного отклонения от типовой последовательности

4. Оперативный контроль введения данных в ЭД, отслеживание отклонения вводимых данных от найденной модели. Анализ допустимости подобной ситуации, проведение уточняющего диалога с пользователем

Для реализации данного прототипа предлагается добавить в ИС встраиваемый интеллектуальный компонент [3], включающий в себя несколько вспомогательных таблиц базы данных, содержащих в себе:

1. Перечень моделей продукционных правил
2. Перечень правил для задания алгоритмов анализа базы данных
3. Вспомогательные таблицы для хранения промежуточных данных (например, таблица с перечнем установленных между таблицами отношений)

4. Перечень возможных отношений в виде функций и логических операторов между таблицами БД или полями одной таблицы

Каждая модель продукционных правил должна состоять из трех частей:

1. Условия вхождения документов в текущую проверку
2. Оператор сравнения данных между моделью и проверяемым ЭД
3. Реакция системы на прохождение проверки

Построим пример модели продукционного правила.

Пример реализации для связи первого порядка (прямой связи между двумя таблицами БД) на основе одного ссылочного поля

Рассмотрим описание правила.

Правило **A**: если поля ЭД $p_1...p_i$ удовлетворяют условиям **P**, то в текущем документе в поле **x** должна присутствовать ссылка на другой ЭД, типа **T**.

Пример построенного на основе модели правила (на языке ABAP (SAP ERP)) [1]:

```
IF table1-field1 = P1.
```

```
  SELECT * FROM table2 WHERE table2-key = x.
```

```
  IF sy-subrc = 0 “Выбор записи из таблицы БД прошел успешно
```

```
  ELSE.”Запись в таблице БД не найдена
```

```
  ENDIF.
```

```
ENDIF.
```

В общем случае, для реализации автоматизированного построения представленного правила предлагается проводить поиск соответствий таблиц моделям правил в следующем порядке:

1. Поиск связей между таблицами БД

После данного шага должны быть выявлены таблицы и/или их поля, между которыми существует взаимоотношение, указанное во вспомогательной таблице 4.

2. Выделение среди найденных связей уникальных параметров

После определения отношений между полями таблицы/таблиц БД должны быть выделены условия, при которых данное отношение существует

3. Определение реакции системы на построенное правило

После проверки полей таблицы/таблиц БД необходимо задать реакцию системы на отклонение данных в текущем ЭД от типовых ЭД (хранящихся в БД)

Для правила **A**:

1. Алгоритм поиска связей представлен на рис. 6.

2. Выделение среди найденных связей параметров предлагается осуществлять с помощью нахождения идентичных полей для заданного процента ЭД, находящихся в одной группе (имеющих идентичное отношение с другой таблицей/таблицами). Т.е. если порог вхождения поля в параметры равно $P\%$, то все поля, значение в которых в пределах одной группы идентично более чем для $P_1\% > P\%$ документов, будут входить в перечень входных параметров текущего правила.

3. Реакция системы определяется на основе $P_1\%$. Если $P_1\% = 100\%$, то при отклонении ЭД от модели необходимо выводить ошибку, запрещающую сохранять документ в БД. Если $75\% < P_1\% < 100\%$, то выводить запрос на подтверждение действий. Если $P_1\% < 75\%$, то выводить предупредительное сообщение. При необходимости реакция системы может быть изменена квалифицированным сотрудником.

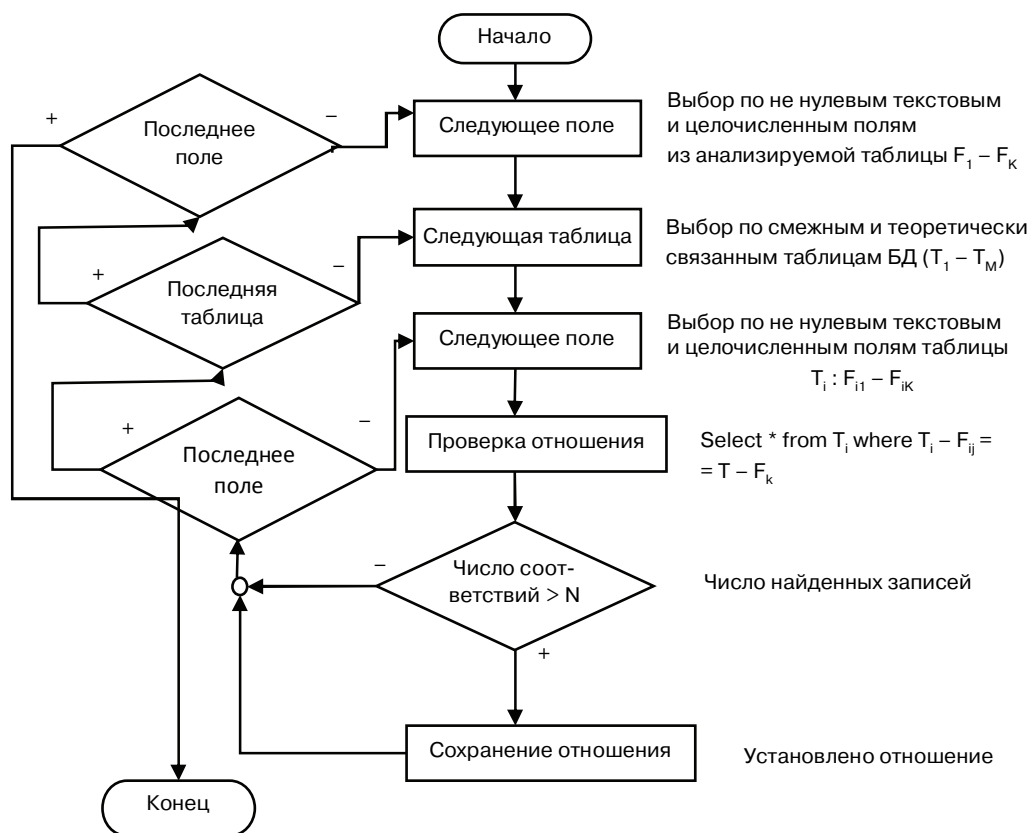


Рис. 6. Алгоритм поиска связей для правила **A**

Для поиска связей между таблицами БД введем следующий набор дополнительных параметров:

1) минимальное число совпадений полей одной таблицы с полями другой таблицы, при котором будем считать, что между таблицами установлено отношение (Для каждой компании определяется экспериментальным путем в зависимости от размеров компании и ее документооборота);

2) типы полей таблиц, между которыми может быть установлено отношение (например, ограничить проведение поиска связей исключительно по целочисленным полям). Обязательность совпадения типов (например, целочисленное поле одной таблицы может ссылаться на текстовое поле другой таблицы);

3) ограничение операторов сравнения (например, поле 1 табл. 1 может всегда частично содержаться в поле 2 табл. 2. В этом случае вместо оператора сравнения необходимо использовать оператор включения, что увеличит время анализа);

4) ограничение круга таблиц поиска $T_1 - T_M$ (например, проводить поиск только по таблицам, связанным с логистическим модулем).

Заключение

Данный подход позволяет повысить мобильность ИС, скорректировать допущенные при внедрении ошибки.

В данной работе выполнен анализ основных принципов по ведению электронного документооборота в крупных компаниях с помощью ИС и, в частности, ERP-систем.

Представлена специфика представления и обработки электронных документов на примере системы SAP ERP.

Предложен прототип, позволяющий решить наиболее часто встречаемые проблемы с помощью применения интеллектуального анализа базы данных продуктивно эксплуатируемой системы.

Предложен алгоритм формирования производственных правил на основе типовых моделей, позволяющий проводить оперативный контроль за вновь создаваемыми электронными документами.

Представленная методология подходит для периодического анализа базы данных и нахождения отклонений в цепочках документов, восстановления консистентности системы, в частности, восстановления корректной хронологии формирования ЭД, обеспечения соответствия данных документа с его моделью.

Разрабатываемый прототип должен позволить внедряющим компаниям сократить затраты на программирование и консалтинговые услуги, снизить риски, связанные с человеческим фактором.

ЛИТЕРАТУРА

- [1] Рюдигер К., Вольфганг В. Разработка приложений SAP R/3 на языке ABAP/4. — Москва, Лори, 1998. [*Rjudiger K., Vol'fgang B. Razrabotka prilozhenij SAP R/3 na jazyke ABAP/4. — Moskva, Lori, 1998.*]
- [2] Бунтова О.Г. Введение в ERP-системы. SAP. Учебное пособие. — Екатеринбург: Уральский госуниверситет, 2007. [*Buntova O.G. Vvedenie v ERP- sistemy. SAP. Uchebnoe posobie. — Ekaterinburg: Ural'skij gosuniversitet, 2007.*]
- [3] Рыбина Г.В. Основы построения интеллектуальных систем: учебное пособие. — Москва, ИНФРА-М, 2010. [*Rybina G.V. Osnovy postroeniya intellektual'nyh sistem: uchebnoe posobie. — Moskva, INFRA-M, 2010.*]
- [4] Dickersbach J.T., Keller G. Production Planning and Control with SAP ERP. — Bonn, Galileo Press, 2010.
- [5] URL: <http://www.erp-online.ru/>

USAGE OF INTELLECTUAL ANALYSIS METHODS IN PROCESSES OF ELECTRONIC DOCUMENTS CIRCULATION

D.O. Grinko

Department of Cybernetic and Mechatronic
Russian People Friendship University
Ordgonikidze str., 3, Moscow, Russia, 115419

This article contains a brief information about the conduct of electronic document circulation (EDC) in companies using information systems (with SAP ERP as example). Includes an analysis of the basic principles of using EDC and disadvantages of standard solutions. An approach is proposed of using intellectual data mining to keep the systems integrity and to control operational data entry. Example of using this approach was proposed.

Key words: electronic document circulation, SAP, information systems, intellectual data mining, document processing in information system, electronic document.