

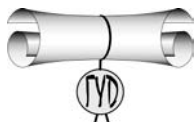
Опыт эффективной организации электронного архива с использованием современных технологий извлечения данных

Полтев Сергей Валерьевич, руководитель направления современных ЕСМ-решений компании «Электронные Офисные Системы» (компания «ЭОС» - член «Гильдии Управляющих Документацией»)

Экспоненциальный рост объемов корпоративных данных, наблюдающийся в последние несколько лет, делает еще более актуальной задачу организацию надежного хранения и эффективного, своевременного и наглядного доступа к архивной информации. Компания ЭОС на протяжении 18 лет реализует проекты по организации электронных архивов самого различного масштаба. Спектр пользователей наших архивных решений – от малых предприятий до многопрофильных холдингов. Среди клиентов – федеральные и региональные органы власти, ведущие производственные компании и финансовые организации. Опираясь на накопленный опыт, и постоянно отслеживая актуальные тенденции, мы можем отметить существенные изменения, и в запросах пользователей архивных систем и в применяемых технологиях для построения электронных архивов технологиях. Растущие объемы данных, распространение планшетных устройств, перспективы постепенно перехода к облачным технологиям – все это требует качественно новых подходов и новых решений в области электронных архивов. О некоторых из этих тенденций мы и постараемся рассказать.

Одно из ключевых требований, обусловленное значительным увеличением объемов, – минимизация времени и трудозатрат на загрузку данных в архив, а также обеспечение просто и удобного представления хранимых данных для пользователей.

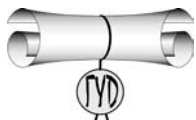
Если речь идет об оцифровке бумажного архива, то, как правило, наиболее трудоемким процессом принято считать сканирование. Но собственно сканирование – это только первый шаг к эффективному и удобному электронному архиву. Для последующей работы с полученными электронными образами каждый из документов необходимо снабдить определенным набором метаданных (заполнить «карточку документа», содержащую определенный перечень реквизитов), – в дальнейшем это позволит использовать различные поисковые возможности и фильтры, создавать различные наглядные представления и разграничивать права доступа для пользователей.



Но если на сканирование документа уходит не более нескольких секунд, то заполнение вручную соответствующей формы, требует как минимум нескольких минут времени. Таким образом, загрузка документов в электронный архив требует значительного времени и существенных трудозатрат, связанных в первую очередь за заполнение карточек документа. Однако подобную ситуацию можно значительно улучшить, если воспользоваться технологиями извлечения данных из неструктурированных документов. С помощью технологии Data Capture можно обеспечить автоматическое заполнение карточки реквизитов данными, извлекаемыми непосредственно из неструктурированного текста документа.

Другой важный момент – соответствие возможностей архивной системы и реальных потребностей по хранению и извлечению данных. Например, соответствие типов документов и их реквизитного состава и необходимых на практике атрибутов для дальнейшего структурирования и обработки информации. Во многих используемых и по сегодняшний день архивных системах возможный набор атрибутов и их визуальное представление либо фиксировано, либо их изменение требует значительных усилий. В результате усложняется поиск необходимой информации, усложняется ее структурирование и подготовка статистических выборок. Такое решение было связано, в том числе с надежностью хранения информации в базе данных и обеспечению достаточной производительности. В то же время, использование современных технологий обработки данных, позволяет обеспечить одновременно надежное хранение документов в базе данных, высокое быстродействие и возможность быстрого конфигурирования атрибутивного состава документа и его визуального представления в точном соответствии со спецификой документа и пожеланиями пользователей.

Фактически, совместное использование гибко-настраиваемой платформы и инструмента для извлечения данных, позволяет выделить в документах значимый набор атрибутов, адаптировать архивную систему в точном соответствии с выбранным атрибутивным составом и его наиболее предпочтительном визуальном представлении и организовать автоматическое заполнение карточек документов при сканировании. Таким образом, обеспечивается полное соответствие хранимых атрибутов документа потребностям пользователей и минимальное время загрузки документов в электронный архив. Опыт внедрений разработанного нами комплексного решения, подтверждает, что применение перечисленных технологий позволяет значительно ускорить как процедуру загрузки информации в электронный архив, так и обеспечить быстрый и удобный доступ к хранимым в архиве данным для пользователей.



Очевидно, что повсеместное увеличение объемов данных продолжится, а следовательно подобные решения, уже востребованные в организациях, деятельность которых связана с постоянной обработкой значительных объемов информации, получат повсеместное распространение. Все большее проникновение облачных сервисов и планшетных устройств будет задавать новый уровень требований к пользовательским интерфейсам в том числе и архивных приложений, и к возможностям «кастомизации» и адаптации внешнего вида приложения и логики его работы в соответствии с потребностями пользователей, а не наоборот как это происходило ранее. Именно на эти тенденции мы стараемся ориентироваться при разработке наших решений для электронных архивов и опыт наших клиентов подтверждает правильность выбранного направления.