

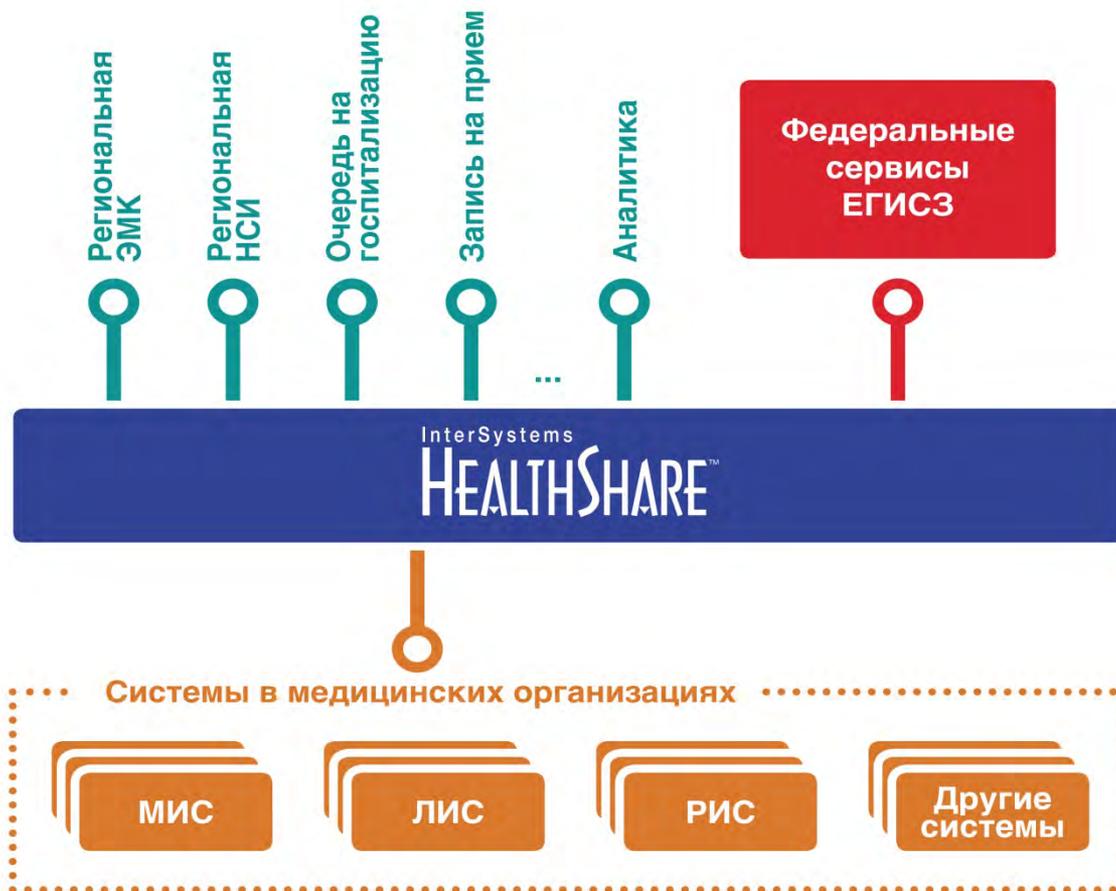
INTERSYSTEMS[®]

Работа над ошибками: региональный мастер-индекс пациентов на платформе InterSystems HealthShare

Дмитрий Засыпкин
InterSystems Russia



Региональный сегмент ЕГИСЗ



Проблемы идентификации пациентов

- Неполные данные
- Ошибки ввода
 - Орфография, опечатки, латинские буквы
- Использование фиктивных значений в обязательных полях
 - СНИЛСы: 123-456-789 00, 111-111-111 45, ...
- Перестановки полей, например, имени и фамилии
- Ввод чужих данных
 - Номер паспорта матери в данных новорожденного ребенка
- Смена фамилии, имени, паспорта, полиса, ...
- Специфические ошибки при вводе номеров документов
- Варианты написания отчеств
- Другие ошибки

Способы идентификации пациентов

- Детерминированные алгоритмы
 - Идентификация пациентов путем применения правил вида
 - «Идентифицировать пациента по первому непустому значению реквизита из списка [СНИЛС, номер полиса ОМС, номер паспорта, ...]»
 - Требуется абсолютная уверенность в полноте и «чистоте» заполнения ключевых реквизитов
- Вероятностные модели
 - Идентификация пациентов с применением вероятностных моделей, фонетических алгоритмов, алгоритмов нестрогого сопоставления
 - Корректная идентификация даже при наличии ошибок и пробелов в данных
 - Низкая производительность по сравнению с детерминированными алгоритмами

Способы идентификации пациентов [продолжение]

- Комбинированный подход
 - Поддержка кросс-таблицы «локальных» идентификаторов пациентов, применяемых в системах-источниках данных
 - Использование вероятностных моделей только в случае отсутствия локального идентификатора в кросс-таблице
 - Использование детерминистских правил для обработки особых случаев

РМП-ID	ЛПУ	Локальный идентификатор пациента	ФИО
1	A	A0815	Скрынник Евгений Петрович
2	A	A1234	Е.Д. Семак
1	A	A4711	Скрыник Е.П.
1	B	B987	Скрынник Е.П.
2	B	B911	Симак Е.Д.
2	C	C22333	Семак Е.Д.
2	C	C22334	Скрынник Е.Д.
3	C	C22335	Скрынник Женя
3	C	C22336	Скрынник Е.Е.

Региональный мастер-индекс пациентов

- РМП на платформе InterSystems HealthShare



- Единый регистр демографической информации
 - Сервисы идентификации пациентов, в том числе по неточным или неполным демографическим данным
 - Интерфейс предметного администратора: список задач на принятие решений о связывании записей
- Возможность использования для вычистки дублей из унаследованных массивов демографических данных

Вероятностная модель

- Используется модель, впервые описанная в статье Fellegi и Sunter “The Theory for Record Linkage” в 1969 году
- Наивный байесовский классификатор (Naïve Bayes) - простой вероятностный классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости
 - Пример: если фрукт красный, круглый и размером около 7см, то есть вероятность, что это яблоко. Применяя наивный байесовский классификатор, мы предполагаем, что эти три свойства фрукта не зависят друг от друга, и наличие каждого из них вносит независимый вклад в вероятность того, что фрукт является яблоком.
- http://en.wikipedia.org/wiki/Naive_Bayes_classifier
- http://en.wikipedia.org/wiki/Record_linkage

Нормализация и индексирование данных

- Специфическая функция нормализации для реквизита
 - Для СНИЛС, например, учитывается наличие контрольной суммы
 - Римские цифры в серии свидетельства о рождении
- Простые преобразования
 - Замена латинских букв и буквы ё, приведение к единому регистру, удаление концевых пробелов
- Удаление null-значений
 - Для каждого из реквизитов может быть настроен собственный список таких значений
- Использование словарей нормализации
 - Юля -> Юлия
- Индексирование: быстрый отбор потенциально совпадающих записей
 - Фонетические алгоритмы для борьбы с опечатками

Вычисление веса пары записей

- Вес пары записей равен сумме весов, полученных при сравнении ключевых реквизитов пациентов с использованием алгоритмов нестрогого сопоставления

	<i>Agree: +10</i> <i>Disagree: -5</i>	<i>Agree: +8</i> <i>Disagree: -8</i>	<i>Agree: +15</i> <i>Disagree: -9</i>	<i>Agree: +6</i> <i>Disagree: -1</i>	<i>Agree: +14</i> <i>Disagree: -9</i>
	ФИО	ДР	СНИЛС	Адрес	Полис ОМС
Источник А	Иванов Дмитрий Иванович	10/12/1956	123-456-789 64	г. Москва, ул. Харьковская д.1, кв. 23	617551217 1234567
Источник В	Иванов Д.И.	нет данных	123-456-789 64	Москва, Харьковская 1-23	234355235 6213232
	+8	0	+15	+5.5	-9

Вес пары: 19.5

Алгоритмы нестроого сопоставления

- `jaroSimilarity`
 - Мера близости Джаро (Jaro similarity measure)
- `winklerSimilarity`
 - Мера близости Джаро-Винклера
- `editDistance`
 - Расстояние Дамерау-Левенштейна
 - Подсчитывается минимальная стоимость набора операций, необходимых для превращения одной строки в другую.
 - Рассматриваются четыре типа операций: вставка одного символа, удаление одного символа, замена одного символа на другой, транспозиция - перестановка двух соседних символов.

Пороговые значения



Интерфейс администратора мастер-индекса

13	<input type="checkbox"/>	Вес связи ▲	Запись	Идентификатор	Имя	Отчество	Фамилия	Дата рождения	Пол	СНИЛС	Просмотр	Действия	<input type="checkbox"/>	13		
1	<input type="checkbox"/>	20.000	A	:44438	Аделаида	Кантемировна	Гакина	01/02/1948	Ж	288-544-152 17	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	1
			B	:99963	Аделаида	Кантемировна	Гайкина	01/02/1948	Ж							
2	<input type="checkbox"/>	20.000	A	:55826		Равшанович	Мусаев	05/03/1954		922-840-515 00	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	2
			B	:78865	Рамиль	Равшан-оглы	Мусаев	05/03/1954	М							
3	<input type="checkbox"/>	20.239	A	:22028	Юлия	Мидхатовна	Гоцева	11/09/2012	Ж	315-376-553 74 123-456-789 00	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	3
			B	:22242	Юлия	Ахатовна	Гоцева	11/08/2012	Ж							
4	<input type="checkbox"/>	20.353	A	:49801	Платон	Маркович	Герциков	02/25/1987	М	779-650-641 68	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	4
			B	:5946	Платон	Маркович	Герциков	02/25/1989	М							
5	<input type="checkbox"/>	20.667	A	:38522	Владимир	Никитич	Симак	09/04/1955	М	442-784-557 10	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	5
			B	:61840	Владимир	Никитич	Семак	09/04/1955	М							
6	<input type="checkbox"/>	21.000	A	:15882	Елена	Ивановна	Никулина	05/27/1988	Ж	123-456-789 00 514-372-225 56	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	6
			B	:57766	Елена	Ивановна	Никулина	05/27/1988	Ж							
7	<input type="checkbox"/>	21.000	A	:25941	Александра	НЕТ	Рашишвили	03/21/2004	Ж	148-964-148 14	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	7
			B	:79167	Александра	Артуровна	Рашишвили	03/21/2004	Ж							
8	<input type="checkbox"/>	22.667	A	:24149	Линара	Григорьевна	Куликова	05/15/1983	Ж	117-030-404 97 117-030-404 97	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	8
			B	:95729	Линара	Григорьевна	Дунаева	05/15/1983	Ж							
9	<input type="checkbox"/>	23.239	A	:26315	Лидия	Федоровна	Титова	01/26/1942	Ж	116-984-038 82	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	9
			B	:89488	Лидия	Петровна	Титова	01/16/1942	Ж							
10	<input type="checkbox"/>	23.539	A	:67643	Анна	Валерьевна	Протопопова	04/06/1947	Ж	890-568-351 54	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	10
			B	:92722	Анна	Валерьевна	Протопопова	04/16/1947	Ж							
11	<input type="checkbox"/>	23.680	A	:32276	Алёна	Владимировна	Боброва	02/01/1977	Ж	801-587-304 89	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	11
			B	:43277	Алена	Владимировна	Боброва	02/01/1977	Ж							
12	<input type="checkbox"/>	23.800	A	:55637	Юлия	Константиновна	Бурмакина	03/04/1948	Ж	756-929-119 57	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	12
			B	:98770	Юля	Константиновна	Бурмакина	03/04/1948	Ж							
13	<input type="checkbox"/>	23.900	A	:28113	Андрй	Ахлиманович	Аникеев	09/21/1928	М	467-058-392 17	Подробности	Связать	Отменить связь	Сбросить	<input type="checkbox"/>	13
			B	:3305	Андрей	Ахлиманович	Аникеев	09/21/1928	М							

Опыт внедрения РМП

- Итерационный процесс настройки параметров мастер-индекса: весов, порогов, алгоритмов, словарей, ...
 - Автоматическая калибровка весов: встроенные алгоритмы MLE, EM
 - MLE - http://en.wikipedia.org/wiki/Maximum_likelihood
 - EM - http://en.wikipedia.org/wiki/Expectation-maximization_algorithm
 - Тонкая подстройка для каждого из источников демографической информации
- Обработка унаследованных массивов данных
 - Зачастую такие данные содержат большой процент дублей
 - Возможность в короткие сроки получить эффект от запуска региональной ЭМК – дать врачам доступ к истории оказания пациенту медицинских услуг
- Наличие сервисов однозначной идентификации пациентов необходимо для эффективного внедрения региональных информационных систем

Сертификаты InterSystems HealthShare

- Сертификат IHE
- Сертификат ФСТЭК

IHE USA 2014 Certification Program
Certificate of Compliance
 Presented this 11th of February 2014 to

InterSystems HealthShare 2014.1

One Memorial Drive • Cambridge, MA 02142 • USA

ICSA Labs is pleased to announce that the product named above attained IHE USA Certification from certification was granted after rigorous and thorough testing to the technical specifications for the following profiles and actors in:

Profile	Actor(s)
Consistent Time (CT)	Time Client
Audit Trail and Node Authentication (ATNA)	Secure Application (SA)
Healthcare Provider Directory (HPD) *	Provider Information Directory
Patient Demographics Query HL7v3 (PDQv3)	PDQv3 Consumer / PDQv3 Supplier
Cross-enterprise Document Reliable Interchange (XDR)	Document Recipient / Document Sender
Cross-Community Access (XCA)	Initiating Gateway / Responding Gateway

*Provisional profile offered for 2014 testing



George P. Japak, Managing Director, ICSA Labs

This product has been certified by the ICSA Labs Certification Body in accordance with the applicable certification criteria set forth by IHE USA.

Certificate number: 140023R00



**СИСТЕМА СЕРТИФИКАЦИИ
 СРЕДСТВ ЗАЩИТЫ ИНФОРМАЦИИ
 ПО ТРЕБОВАНИЯМ БЕЗОПАСНОСТИ ИНФОРМАЦИИ
 № РОСС RU.0001.01БИ00**

**СЕРТИФИКАТ СООТВЕТСТВИЯ
 № 3264**

Выдан 7 ноября 2014 г.
 Действителен до 7 ноября 2017 г.

Настоящий сертификат удостоверяет, что Комплекс средств защиты Платформы создания медицинских информационных сетей HealthShare версии 2014.1 (партия из 100 (ста) экземпляров продукции с заводскими знаками соответствия с № И 201003-1000158-20142551-11-000001 по № И 459200) производства ЗАО «АСТ», функционирующий в средах операционных систем, указанных в формуляре ХЛФШ.502211-1413001, являющийся средством защиты от несанкционированного доступа к информации, не содержащей сведений, составляющие государственную тайну, реализующим функции идентификации и аутентификации субъектов доступа и объектов доступа, управления доступом субъектов информации, обеспечения доступности информации, обеспечения целостности информации, обеспечения конфиденциальности информации, обеспечения безопасности системы, ее средств, систем связи и передачи данных, соответствующим требованиям государственного документа «Защита от несанкционированного доступа к информации и защиты информации» (Гостехкомиссия России, 1999) – по 4 уровню контроля и технических условий ХЛФШ.502211-1419101 при выполнении ограничений по применению, указанных в формуляре.

Сертификат выдан на основании результатов сертификационных испытаний, проведенных испытательной лабораторией ФАУ «ГНИИ ПТЗИ ФСТЭК России» (аттестат аккредитации от 26.04.2005 № СЗИ RU.840.A.92.007) – технического заключения от 12.09.2014, и экспертного заключения от 25.09.2014 органа по сертификации ООО «Центр безопасности информации» (аттестат аккредитации от 30.03.2006 № СЗИ RU.117.B08.025).

Заявитель:
 Адрес:
 Телефон:

ЗАО «АСТ»
 109316, г. Москва, проезд Остاپовский, д. 3, стр. 26
 (495) 679-8686

Аккредитация на осуществление сертификации в соответствии с требованиями руководящего документа и технических условий, выданных в соответствии с требованиями руководящего документа и технических условий, осуществляется испытательной лабораторией ФАУ «ГНИИ ПТЗИ ФСТЭК РОССИИ»

Спасибо за внимание!

- Дмитрий Засыпкин
dmitry.zasypkin@intersystems.com
<http://www.InterSystems.ru>