spellabs it.company

Обзор применения Data Mining с учетом специфики HR-организаций

Михаил Сумской
системный архитектор

Microsoft Partner
Silver Independent Software Vendor (ISV)

spellabs it.company

О компании

Компания spellabs работает с 2004 года
Основные интересы и компетенции:

- Разработка корпоративных портальных решений
- Внедрение систем и методологии анализа данных

Microsoft Partner
Silver Independent Software Vendor (ISV)

spellabs it.company

План доклада

- Data Mining: общее понятие
- Задачи Data Mining и обзор алгоритмов
- Сценарий: выявление факторов влияния
- Сценарий: исследование навигации на сайте

spellabs it.company

Data Mining: общее понятие

Data Mining

Data Mining – это процесс анализа данных с целью выявления в них скрытых закономерностей с помощью автоматических методик.

spellabs it.company

Применение

- Выдача рекомендаций
- Выявление аномалий
- Анализ оттока клиентов
- Управление рисками
- Сегментация клиентов
- Целевая реклама
- Прогнозирование

Задачи Data Mining

spellabs it.company

Классы задач

- Описательный анализ
  - Профиль идеального соискателя
  - Анализ закономерностей карьерных лестниц
  - Взаимосвязь информации в резюме
- Предиктивный анализ
  - Анализ рисков при приеме на работу
  - Прогнозирование спроса на вакансии
  - Предсказание вакансий, подходящих соискателю

spellabs it.company

Классификация

- Откликнется ли соискатель на вашу вакансию?
- Что характерно для соискателя, откликающегося на определенные группы вакансий?

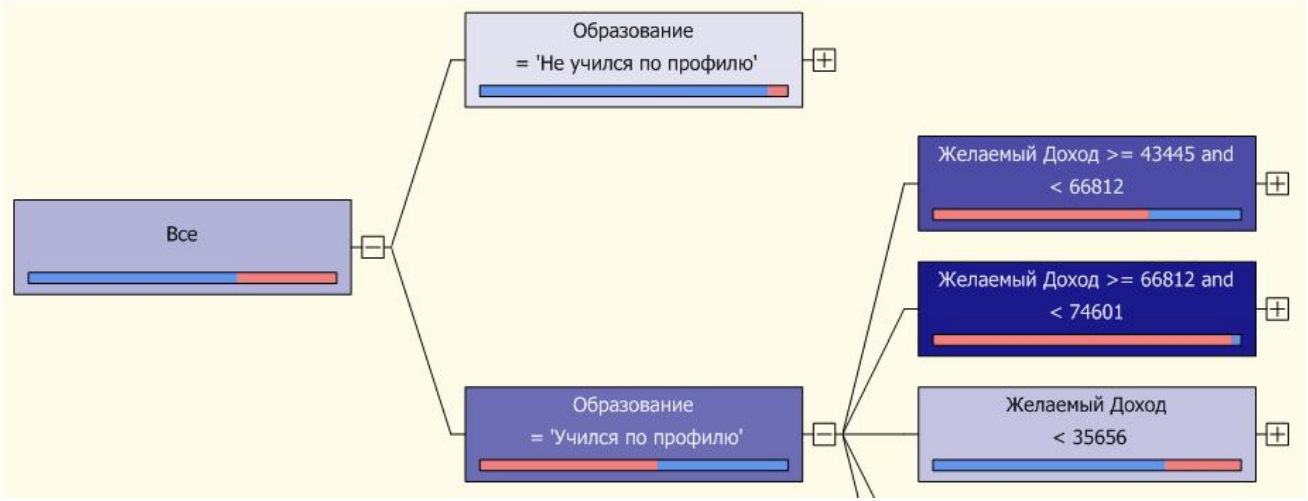


Рис. 1. Анализ желаемого дохода, и образования показал, что если человек не имеет профильного образования, то, скорее всего, он не пойдет работать программистом, а те, кто пойдут – захотят зарплату от 66 до 74 тысяч рублей.

Использован алгоритм Microsoft Decision Trees.

spellabs it.company

Сегментация

- Выявление особенностей естественных группировок резюме, вакансий, соискателей
- Характеристика группировок невостребованных резюме и соискателей
- Выявление скрытых, но репрезентативных групп пользователей

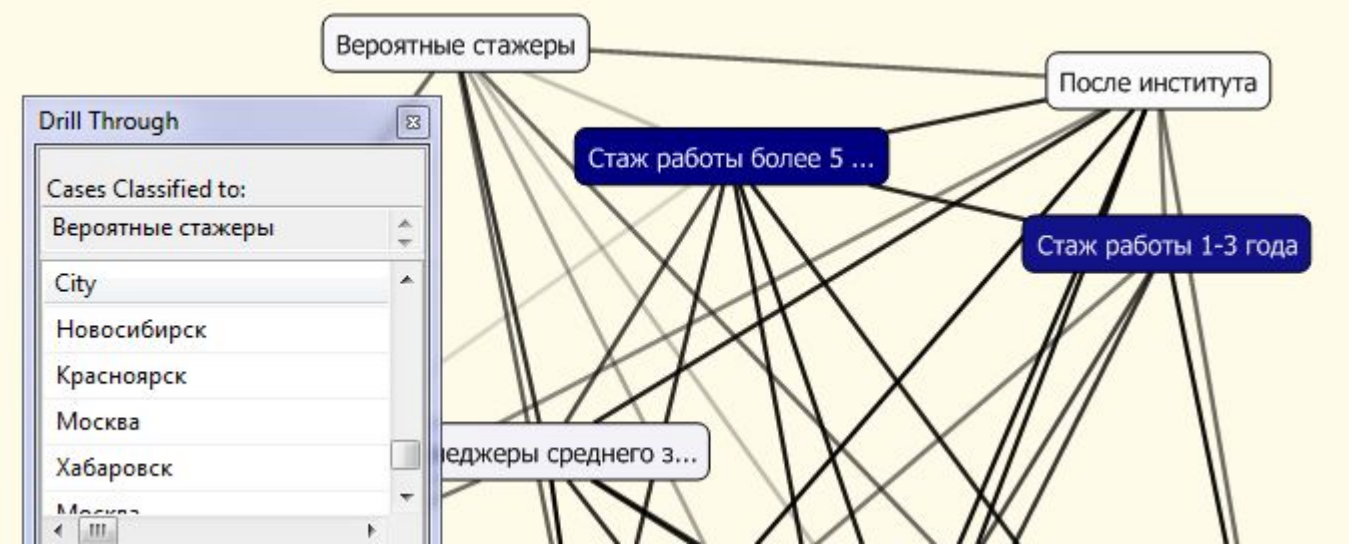


Рис.2. Анализ кластеров показал, что в данной отрасли имеется нехватка молодых специалистов, а москвичи совсем не склонны идти стажерами.

Применен алгоритм Microsoft Clustering.

Microsoft Partner
Silver Independent Software Vendor (ISV)

spellabs it.company

Анализ путей влияния

- Влияние семейного положения на выбор профессии
- Связь между образованием, доходом, и местом проживания

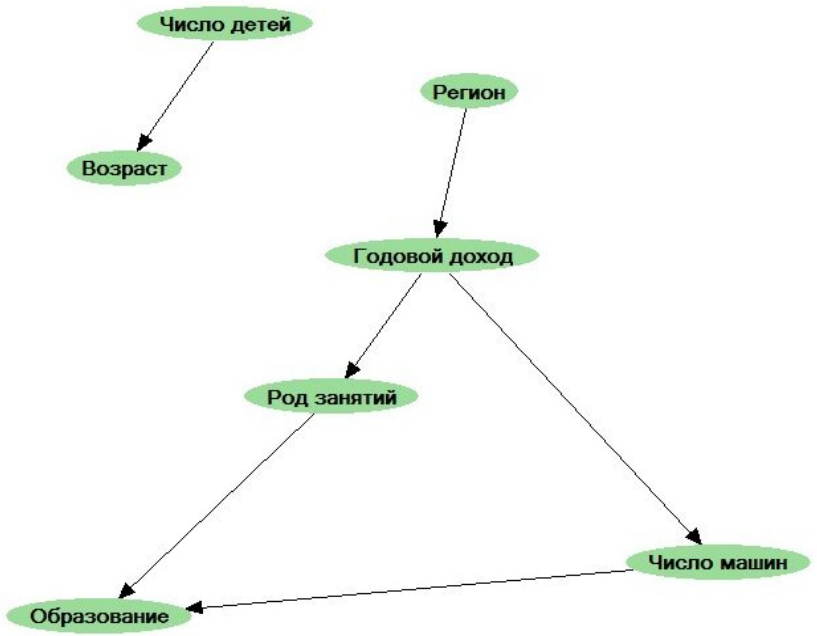


Рис. 3. Граф взаимосвязей характеристик соискателя.

Использован алгоритм Microsoft Naïve Bayes.

spellabs it.company

Прогнозирование

- Прогноз спроса на специалистов
- Прогноз с учетом сезонности
- Прогнозирование динамики рынка вакансий с учетом его сегментов и взаимосвязей с другими отра

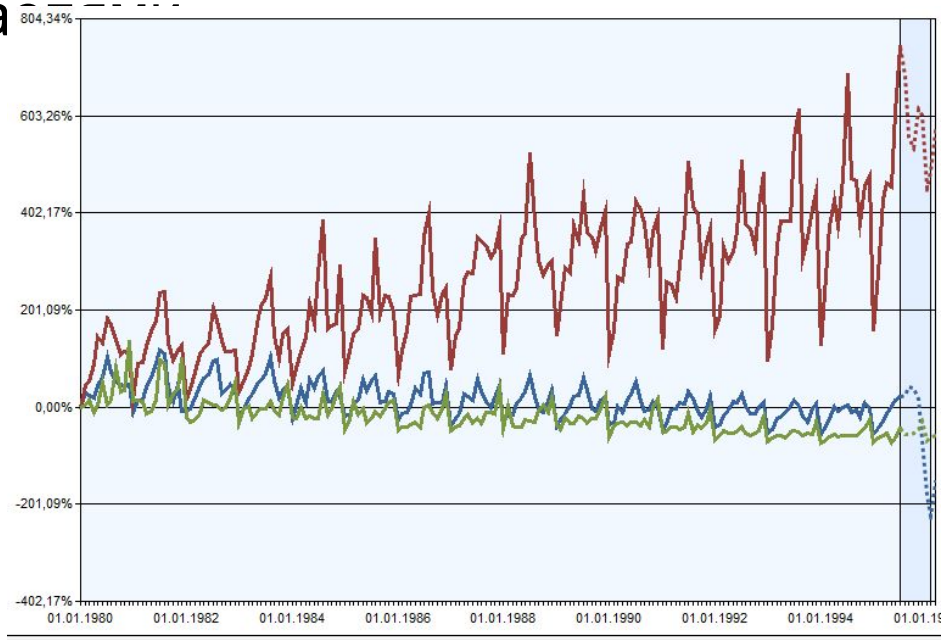


Рис.4. Анализ особенностей отрасли позволил предсказать динамику спроса на программистов на языках высокого уровня с учетом динамики спроса на программистов на двух видах ассемблера.

Использован алгоритм Microsoft Time Series.

Microsoft Partner
Silver Independent Software Vendor (ISV)

spellabs it.company

Ассоциативные правила

- Выявление шаблонов карьерной лестницы
- Каковы наборы предпочитаемых работодателей у начинающих специалистов различных отраслей?
- Рекомендации на основе имеющегося опыта работы и информации из резюме

Рис. 5. Анализ выявил тенденцию, что для соискателей с низким желаемым доходом не характерно желание стать программистами, при этом это решение не зависит от пола, но зависит от образования.

Применен алгоритм Microsoft Association Rules.

spellabs it.company

Анализ цепочек последовательностей

- Какова вероятность ухода с сайта после просмотра данной вакансии?
- Куда пойдет соискатель после просмотра страницы компании?
- Какие сочетания страниц наиболее популярны для данного типа соискателей?

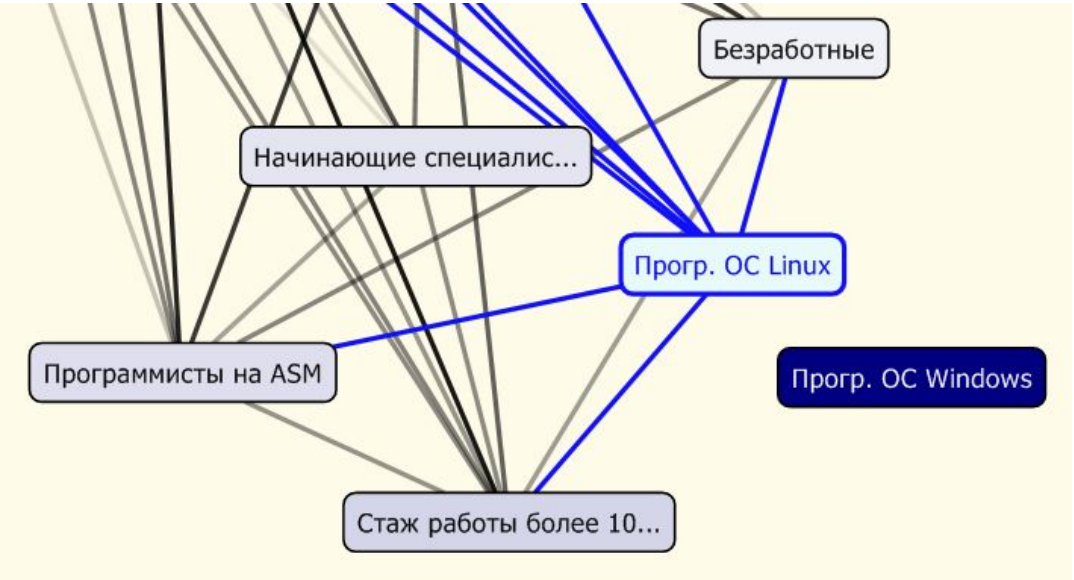


Рис.6. Анализ цепочек переходов на сайте неожиданно показал, что поведенческие мотивы программистов под Windows преобладают на сайте, и сильно отличаются от поведения других программистов, которые “растворяются” среди других категорий пользователей.

Применен алгоритм Microsoft Sequence Clustering.

spellabs it.company

Сценарий: выявление факторов влияния

Microsoft Partner
Silver Independent Software Vendor (ISV)

Особенности сценария

- Необходимость выявления взаимосвязей факторов
- Визуализация в виде ациклического графа
- Требуется независимость модели от количества факторов
- Высокие требования к быстродействию

spellabs it.company

Решение:
spellabs influence.maps

Рис. 7. Анализ анкет американских обывателей с помощью данного решения показал, что со времен одноэтажной Америки кое-что изменилось.

spellabs it.company

Преимущества решения

- Автоматическое выявление факторов влияния
- Сортировка факторов влияния по силе связи
- Возможность ручной корректировки выявленных факторов и пересчета модели с учетом внесенных изменений
- Полная реализация Байесовских сетей
- Визуализация реализована на HTML5

spellabs it.company

Сценарий: исследование навигации на сайте

Microsoft Partner
Silver Independent Software Vendor (ISV)

spellabs it.company

Особенности решения

- Выявление поведенческих шаблонов на сайте
- Выявление частых сочетаний посещенных страниц в рамках пользовательских сессий
- Кластеризация посетителей сайта
- Высокие требования к быстродействию, возможность выполнения предсказания “на лету”

spellabs it.company

Решение: spellabs web.usage mining

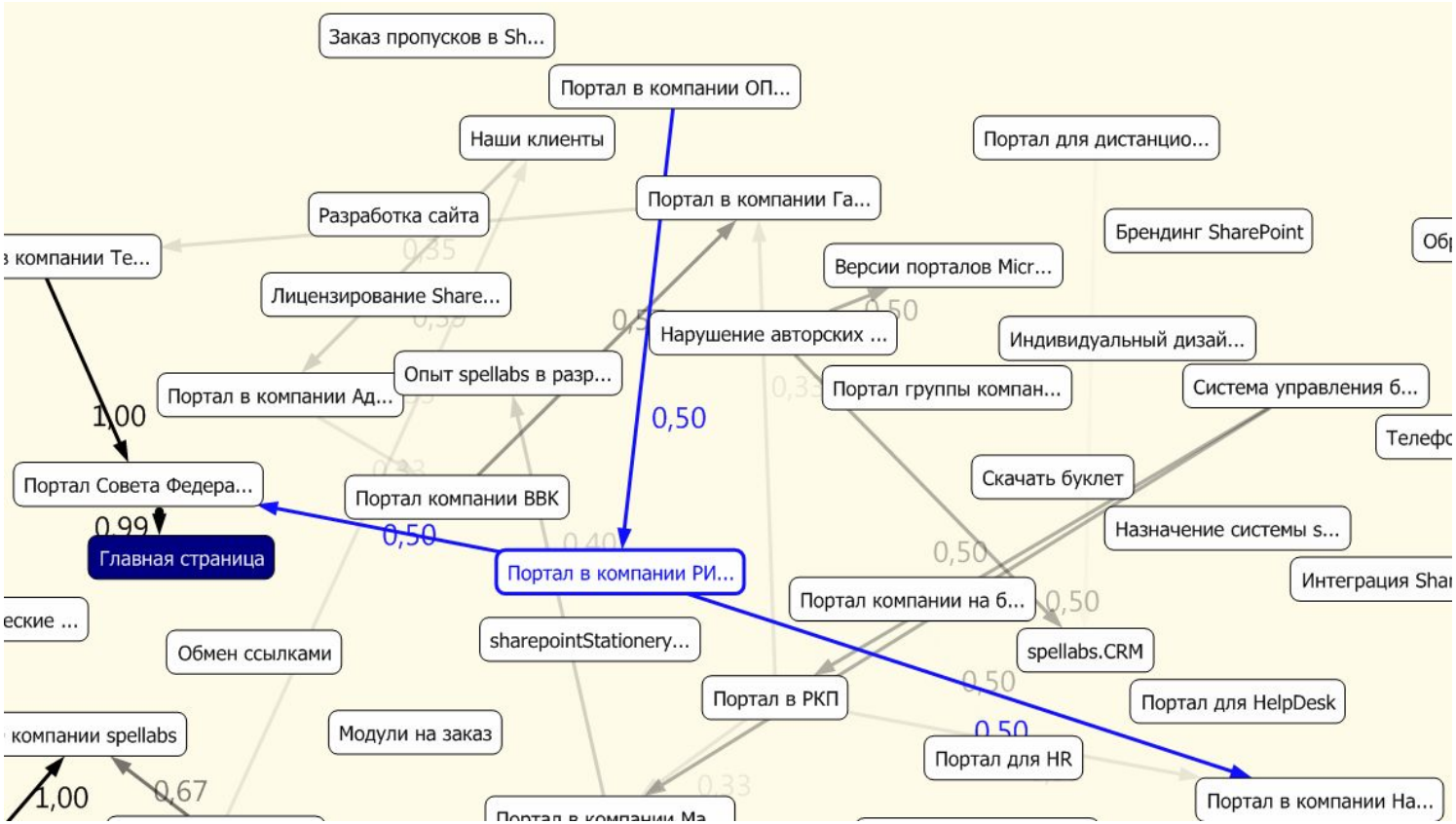
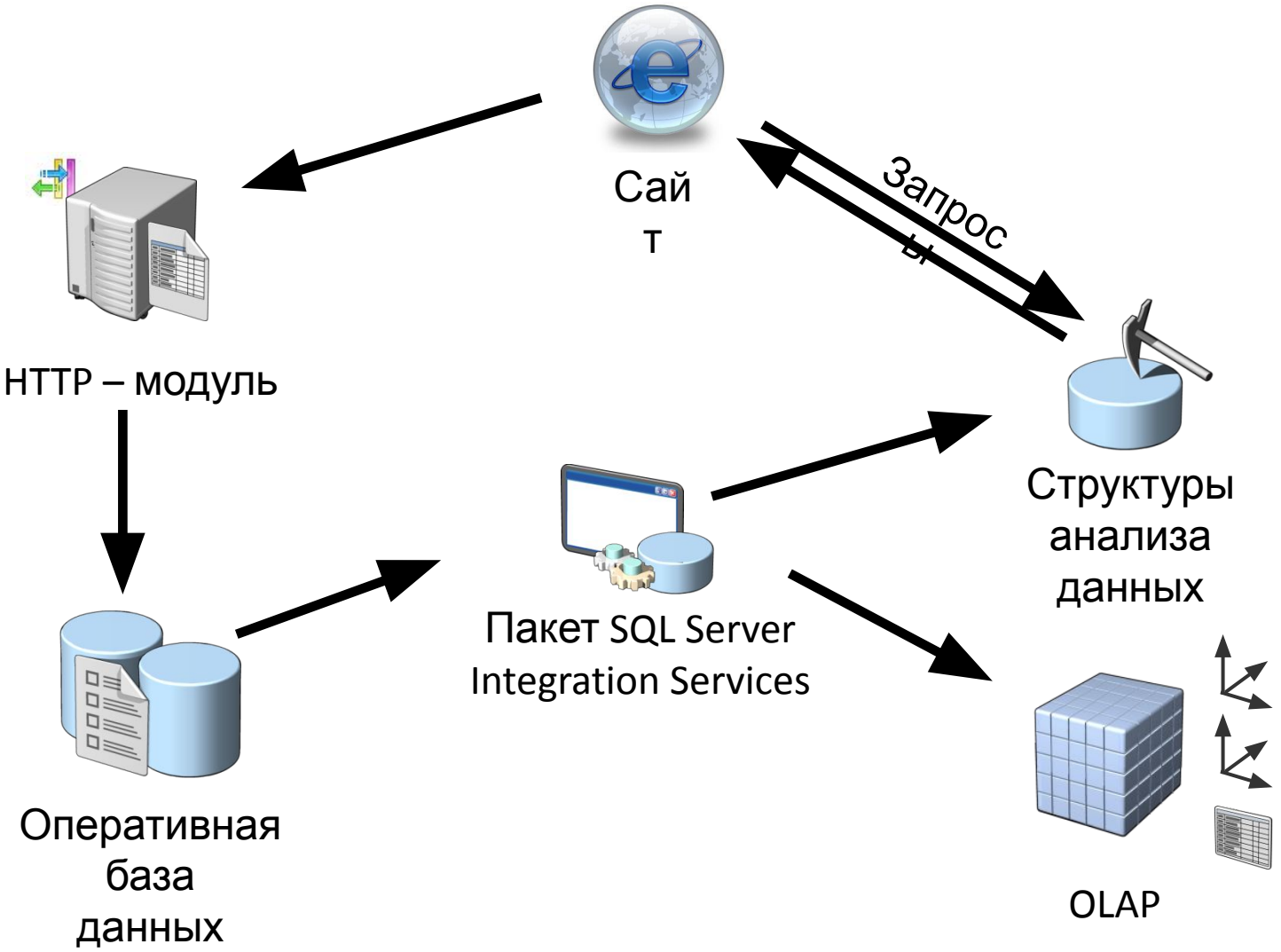


Рис. 8. Просмотр графа посещаемости внутри кластера посетителей сайта spellabs.ru, с вероятностями переходов на другие страницы.

spellabs it.company

Архитектура решения



spellabs it.company

Преимущества решения

- Возможность прогнозирования переходов в зависимости от поведения пользователя
- Быстродействие предсказания
- Выявление “проблемных” страниц, после которых, например, посетитель уходит с сайта
- Кластер пользователя определяется на основе его поведения, возможен учет персональной информации
- Интегрированный в решение OLAP, позволяющий получить представление о посещениях страниц

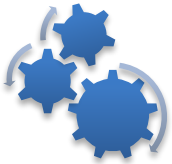
spellabs it.company

Ответы на вопросы

Microsoft Partner
Silver Independent Software Vendor (ISV)

spellabs it.company

spellabs



Спасибо

http://www.businessdataanalytics.ru

актуальные материалы об алгоритмах и технологиях
добычи знаний и интеллектуального анализа данных

http://www.spellabs.ru

сайт нашей компании

Microsoft Partner
Silver Independent Software Vendor (ISV)