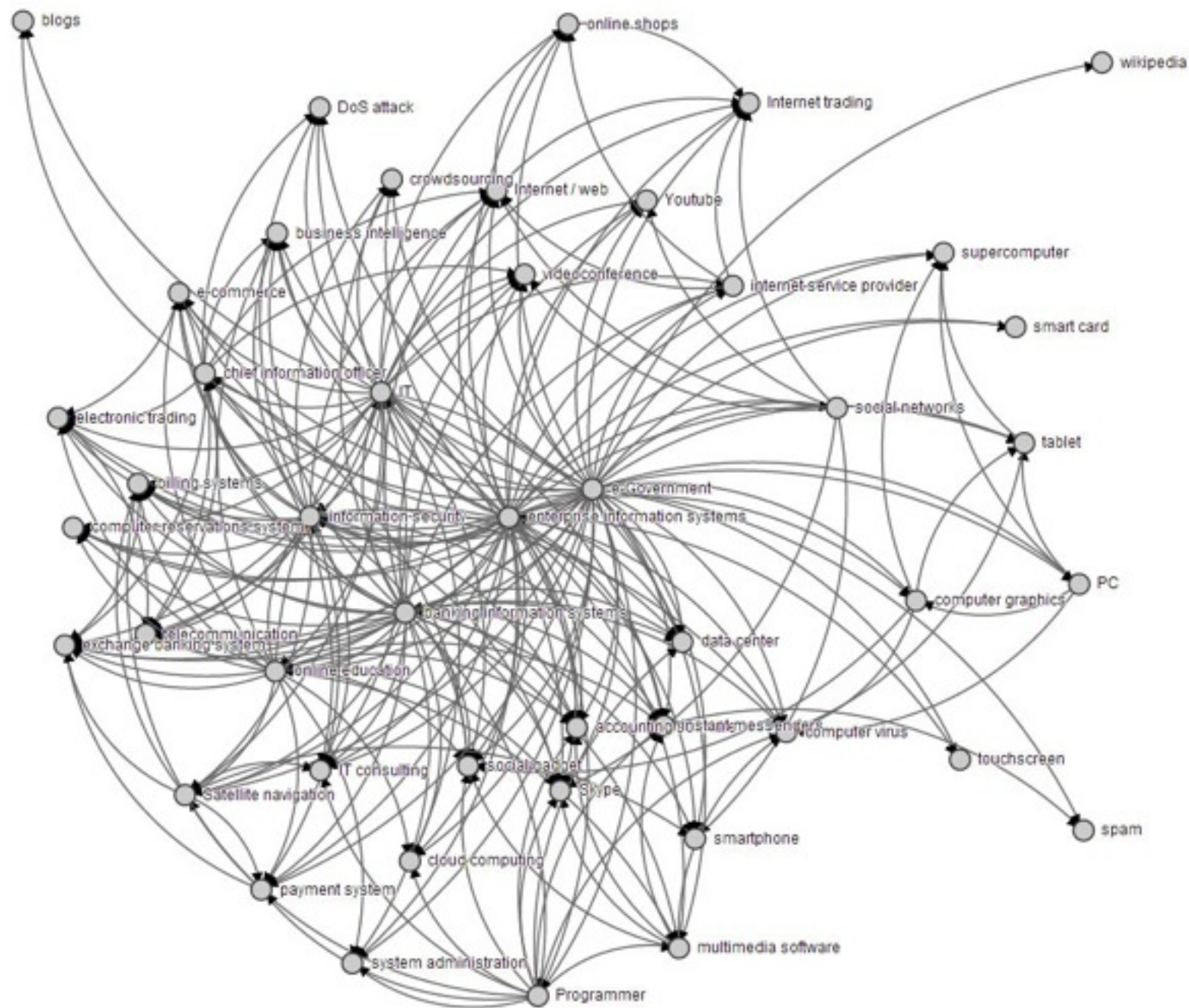


Концептуальные карты: метод построения и анализ

Егор Моренко, **Екатерина Черняк**, Борис Миркин
НИУ ВШЭ

Пример концептуальной карты





Мотивация

Что в больших
коллекциях
текстовых документов?

Входные данные

- Ключевые слова / словосочетания
- Коллекция текстовых документов

2 концептуальные карты “Бизнес”

- Ключевые словосочетания заданы экспертами

Раскрытие ошибок отчетности

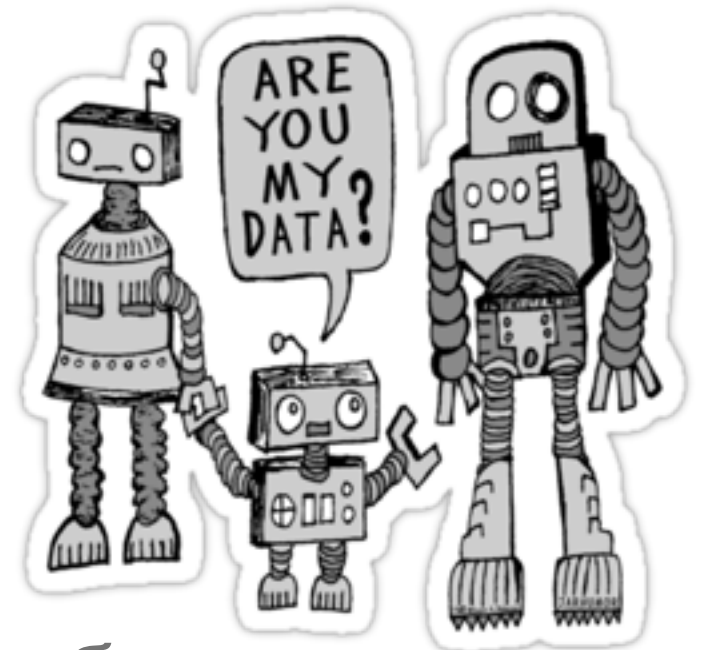
Смена генерального директора

Участие в судебных разбирательствах

Выход на IPO

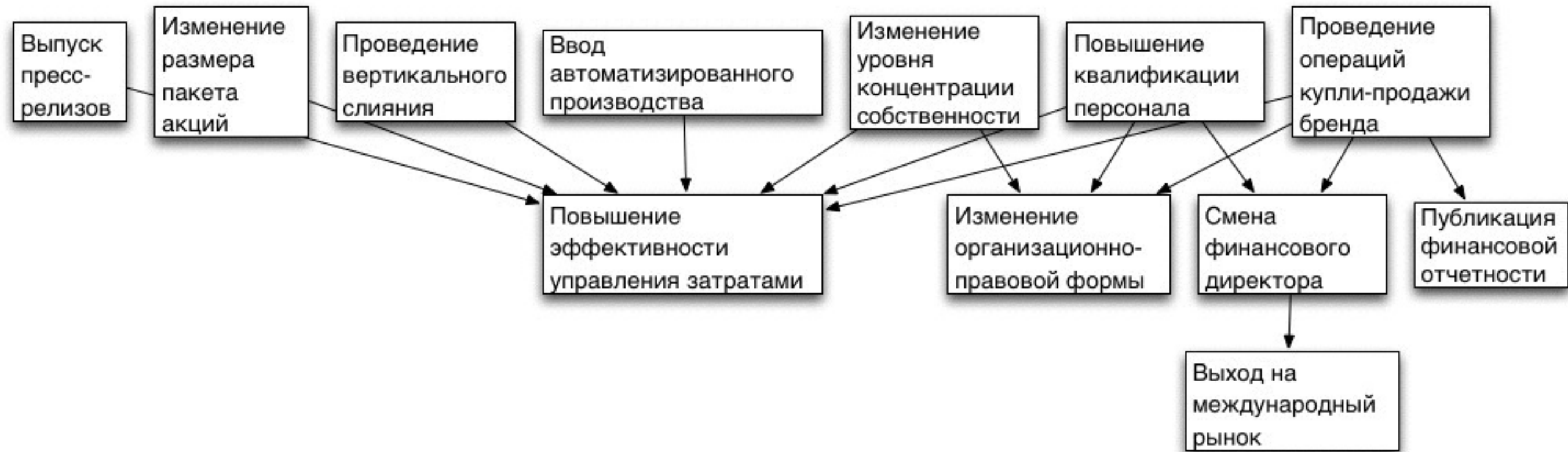
Продажа активов

Сужение бизнеса

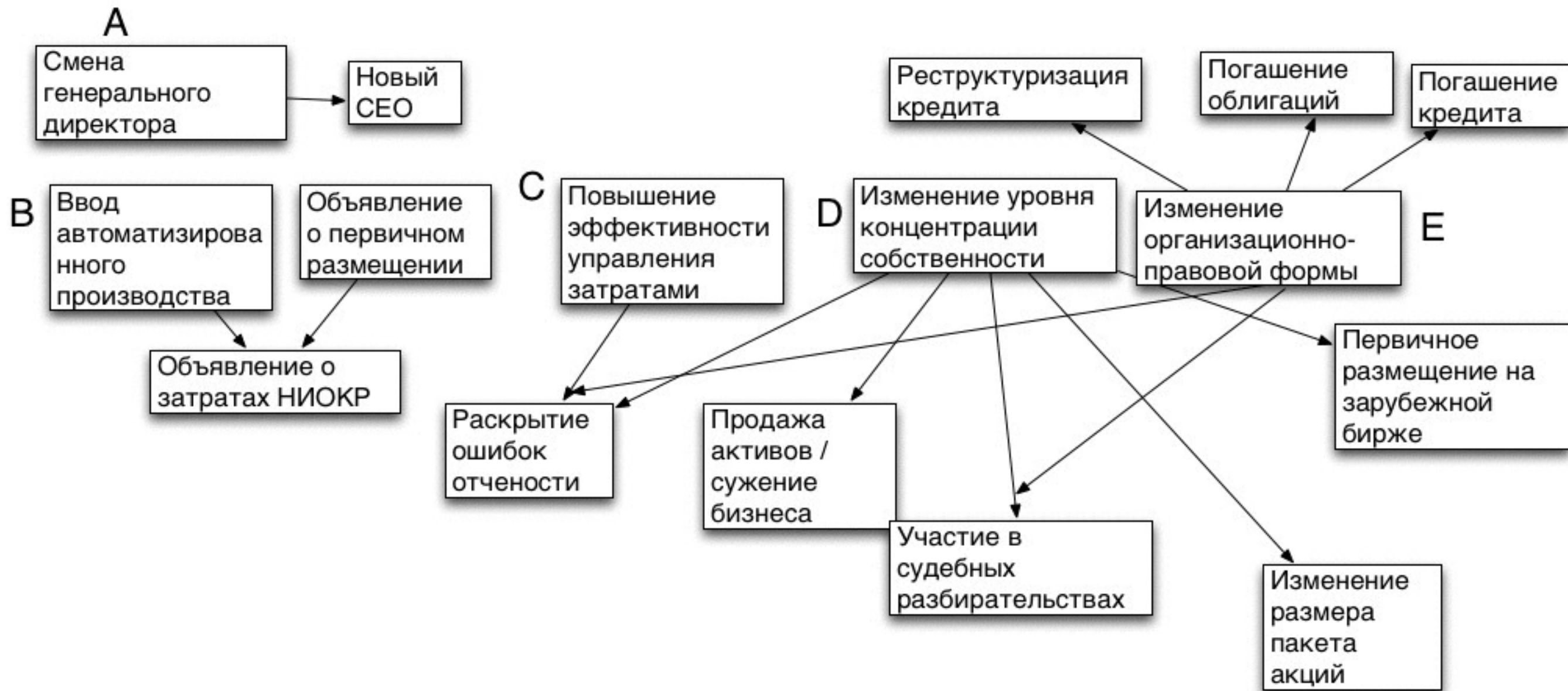


- Коллекция новостных сообщений, опубликованных на порталах газет (~1000 текстов)
- Статьи из Википедии, принадлежащие к категории “Бизнес” (~12000 текстов)

Газетная концептуальная карта



Энциклопедическая концептуальная карта



Метод построения концептуальной карты – 1

- Оцениваем релевантность каждого ключевого словосочетания каждому тексту с помощью меры, основанной на аннотированных суффиксных деревьях
- Определяем для каждого ключевого словосочетания K_i множество текстов $F(K_i)$, которым оно релевантно
- Порог релевантности: r (из интервала $[0.5, 1]$)

Метод построения концептуальной карты – 2

- $K_i \rightarrow K_j$, если

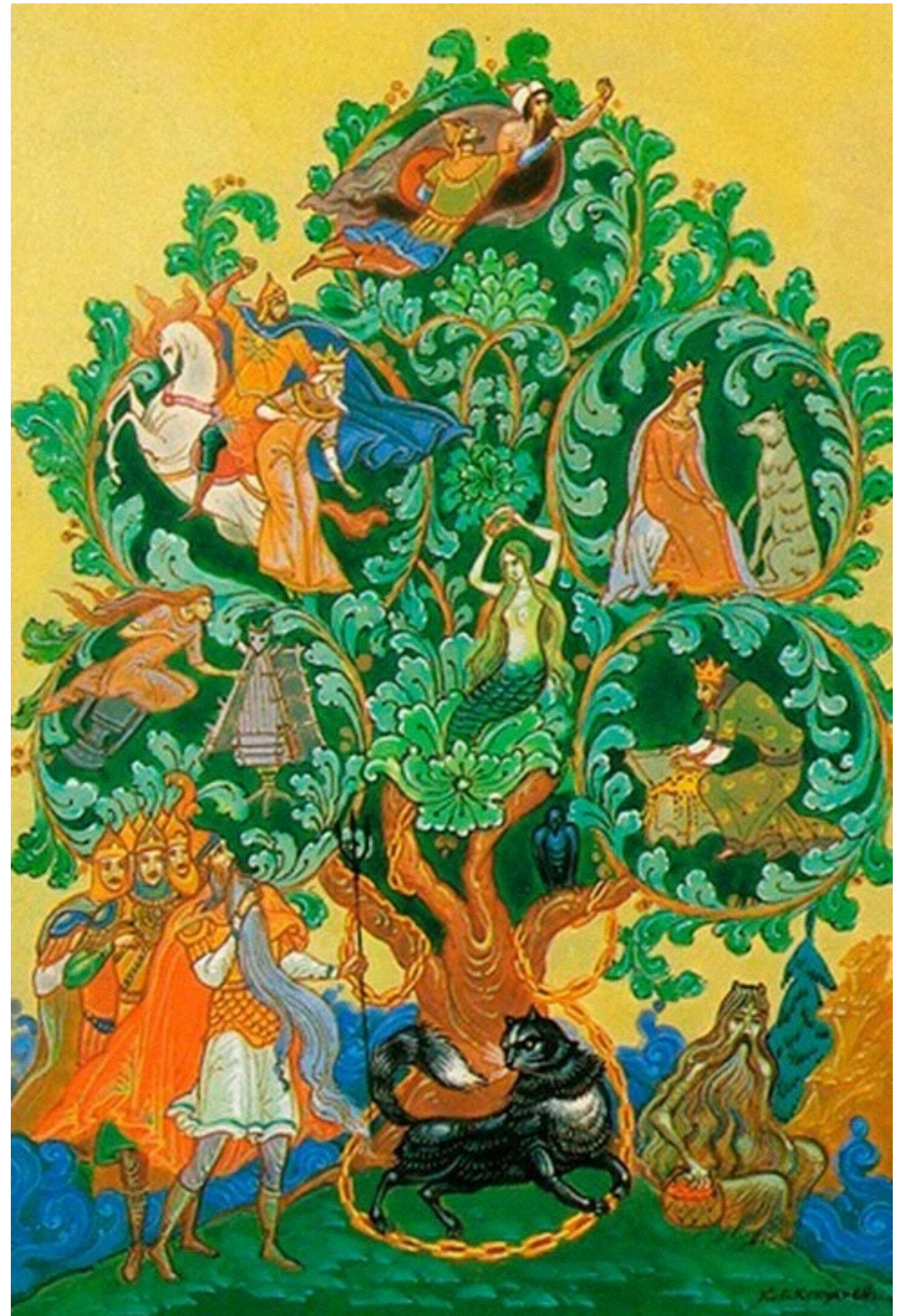
$$\frac{|F(K_i) \cap F(K_j)|}{|F(K_i)|} > c$$

- c – порог, $c \in (0.5, 1]$

- **Пример:** Ввод автоматизированного производства  Повышение эффективности управления затратами

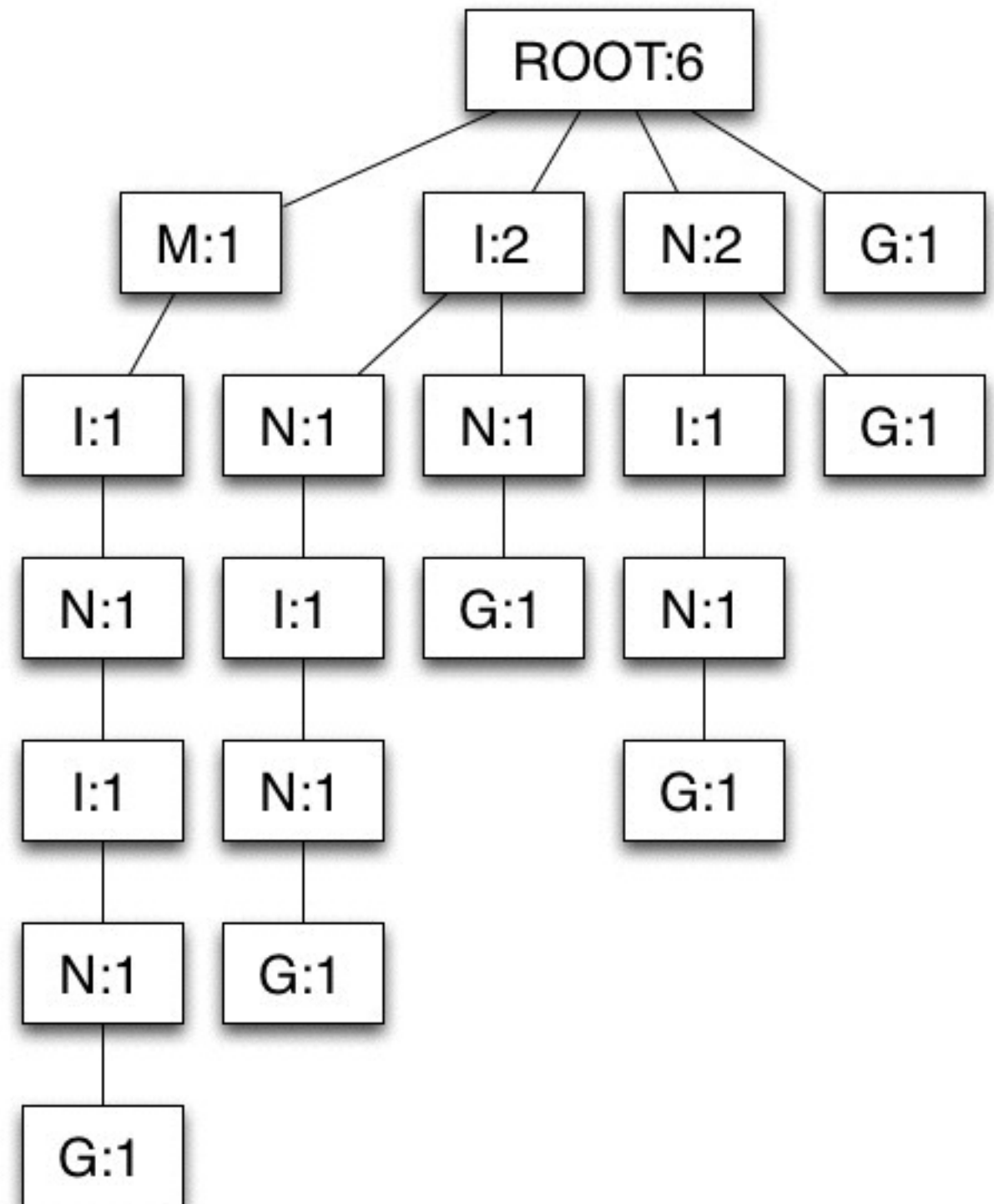
Мера АСД

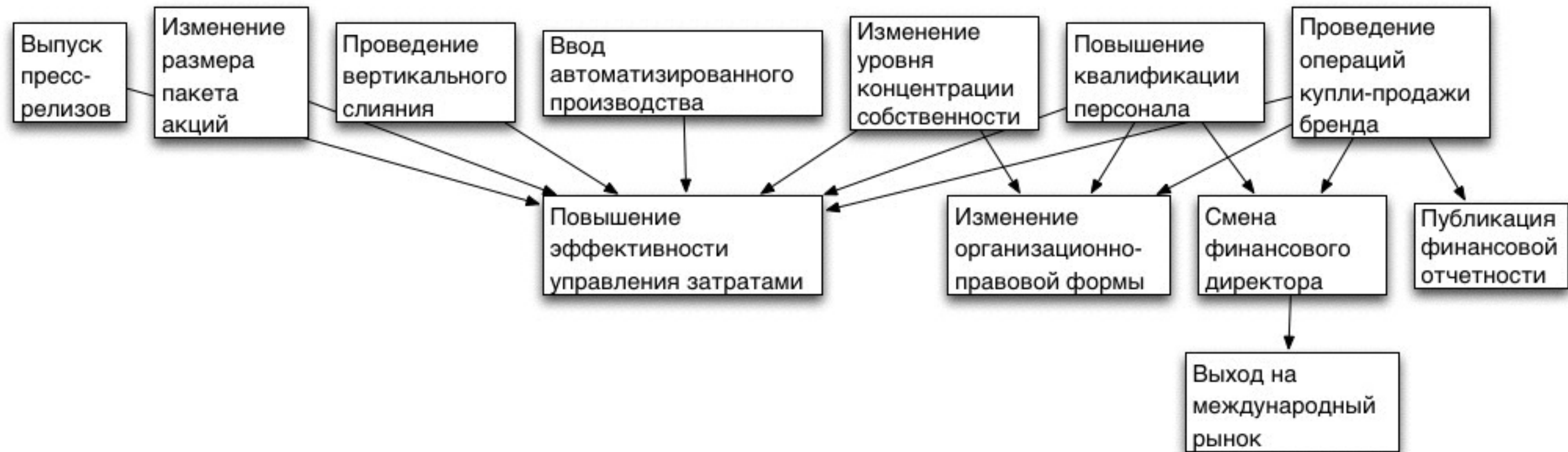
- Ставит ключевому словосочетанию и тексту в соответствие число (оценку релевантности)
- Которое изменяется от $[0,1]$
- Учитывает нечеткие совпадения



Автоматизация производства

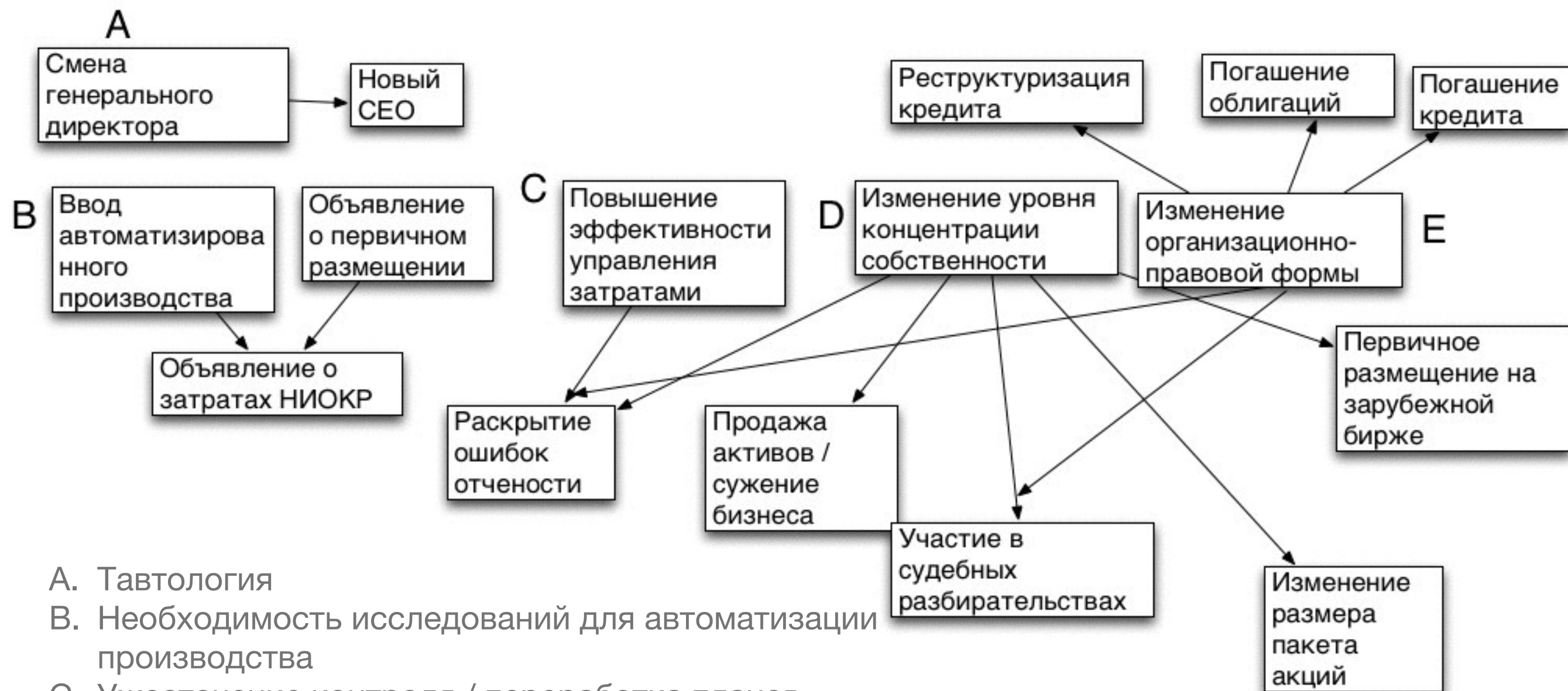
- **Автоматизация производства** — это процесс в развитии машинного **производства**, при котором функции управления и контроля, ранее выполнявшиеся человеком, передаются приборам и **автоматическим** устройствам. Введение **автоматизации** на **производстве** позволяет значительно повысить **производительность** труда и качество выпускаемой продукции, сократить долю рабочих, занятых в различных сферах **производства**. До внедрения средств **автоматизации** замещение физического труда происходило посредством механизации основных и вспомогательных операций **производственного** процесса. Интеллектуальный труд долгое время оставался не механизированным (ручным). В настоящее время операции физического и интеллектуального труда, поддающиеся формализации, становятся объектом механизации и **автоматизации**.





Интерпретация газетной концептуальной карты

- Логическая связь определяется отношением “источник-цель”
- Концептуальная карта передает действующие в бизнесе схемы



- А. Тавтология
- В. Необходимость исследований для автоматизации производства
- С. Ужесточение контроля / переработка планов развития
- Д. Необходимые действия для изменения уровня концентрации
- Е. Необходимые действия для изменения организационно-правовой формы

Интерпретация энциклопедической концептуальной карты

- Логическая связь определяется отношениями “действие-последствие” и “цель-действие”
- Концептуальная карта носит информационный характер

Обзор работы

- Реализован метод построения концептуальных карт
- Построены концептуальные карты трех областей (Бизнес, ИТ, Маркетинг)
- Предложена интерпретация двух концептуальных карт “Бизнес”
- Дополнительно: утилиты для извлечения новостных сообщений, статей из Википедии, обработки текстов

Дальнейшее развитие

- Автоматизация выделения ключевых словосочетаний
- Использование методов теории графов для анализа концептуальных карт
- Расширение на случай правил вида многие-к-одному, многие-ко-многим, один-к-многим



Вопросы?

Спасибо!