

Яндекс

Практика. On- the-fly генерация SDCH-словарей

Алексей Баранов

Яндекс.Браузер. Группа сетевых технологий

Всемирная паутина обладает
низкой энтропией

Мы все хотим загружать страницы быстрее

Уменьшение объема загруженного \rightarrow увеличение энтропии пересылаемого сообщения.

- Кеширование (на клиенте и middleware)
- gzip
- single-page applications
- ... мобильные приложения – тоже способ максимизировать энтропию

Meet SDCH (Shared Dictionary Compression over HTTP)



SDCH?

TL;DR

- 1) Берем много похожих ресурсов нашего сайта, строим общий словарь.
- 2) Посылаем на клиента
- 3) все запрашиваемые странички сжимаем относительно него с помощью VCDiff ([RFC3284](#))
- 4) Profit

Прокси сервер

берем много похожих ресурсов с какого-то сайта, строим общий словарь.

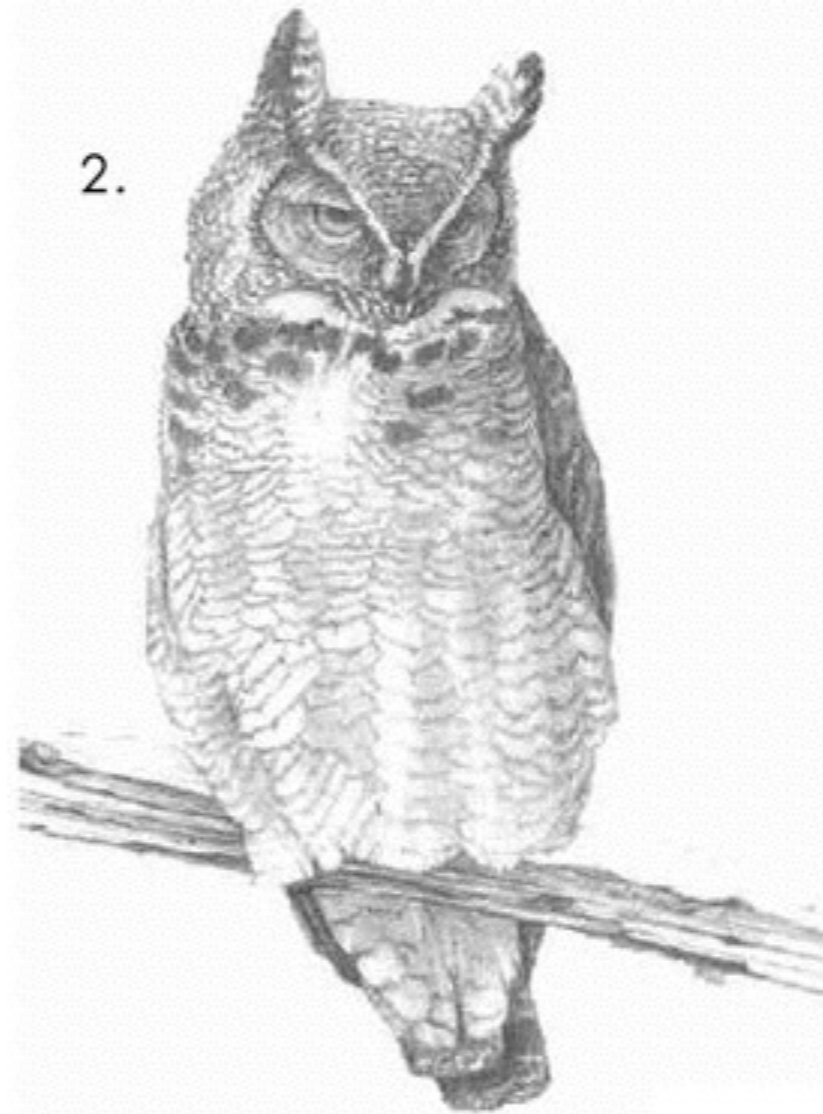
How to draw an owl

1.



1. Draw some circles

2.



2. Draw the rest of the fucking owl

Нет (хороших) инструментов для генерации словарей.

У github.com/gtoubassi/femtozip есть 3 фатальных недостатка.

Что надо будет делать?

Задача:

- создание инструмента для автоматической инкрементальной генерации SDCH-словарей на основе данных, проходящих через HTTP-прокси сервер.
- реализация тестового прокси-сервера на основе этого

Что интересного:

- актуальная сложная нерешенная задача
- работа с performance-critical кодом
- свобода в выборе пути решения

Требования:



- отличное умение Гуглить
- C++
- желательно javascript + nodejs
- знание простой математики, алгоритмов/структур данных ИТД.

Спасибо за внимание!

Алексей Баранов baranovich@yandex-team.ru

Яндекс.Браузер. Группа сетевых технологий