HCE project's team briefly

This document describes briefly a common skills and features of HCE team participants.

Frameworks, languages and libraries

C++: STL, POCO, msgpack, curl, libev, libxml2, zmq, Apache Thrift, tydylib, ImageMagic, pthread, pprocess, iconv, mysqlc, sphinx search API;

Python: cement, Flask, SQLAlchemy, scrapy, gmpy, pyzmq, lepl, requests, urlnorm, pyicu, newspaper, goose-extractor, pytidylib, uritools, python-magic, feedparser, pillow, Ghost, libffi, libxml2, libxslt1, selenium;

PHP: phpzmq, mbstring, gd, mysqli, Yii, Laravel, WordPress;

Java: Android Development Tools;

JavaScript: JQuery, highcharts, firebug, chrome developer tools;

MS Visual Studio: C#, Visual Basic, CLI C++, console and form applications, win32 API, MFC, BHO API.

Environment

Debian Linux, OpenVZ, bash;

Centos Linux, Fedora Linux;

Android;

Windows;

VMWare;

OpenVZ Virtuozzo Containers;

LXC containers;

Chrome browser;

Technologies and tools

MySQL, PostgreSQL, sqlite, Apache, nginx;

gdb, strace, apt, make, build-essential, cppcheck, Cppunit, GoogleTest, pylint, phpdoc, doxygen;

Sphinx search;

TCP, HTTP, JSON, XML, HTML;

Algorithms and techniques

Ternary tree, binary tree, multi-thread data processing, multi-process data processing, MOM transports, network messages distribution patterns, distributed data balancing, distributed data sharding, system resources usage optimization, fast access memory structures, indexed search, web scraping, web crawling, stable Linux daemons optimizations, universal network transports patterns and algorithms, asynchronous sockets, asynchronous message processing, distributed storages, parallel data processing, parallel networking, tasks management, load balancing, REST services, TCP services, serialization protocols and formats, asynchronous sockets pooling;

Dynamic rendered web-site's content scraping, unstructured data scraping and pre-processing with formalization of a structure.

Statistical data analysis including Bayesian statistical implementations of a similarity rate of a documents, sentiment, classification, clusterization, pop-words and related chains detection and similar rates and indicators computations.

Development process and management

The agile SCRUM methodology.

Jira, Mantis, Redmine, oDesk, basecamphq, github;

SVN, Git;

Jenkins;

Eclipse: CDT, PDT, ADT;

KDevelop;

AptanaStudio;

Enterprise Architect, Visual Paradigm, Umbrello, MS Visio;

Skills and experience

Oldest team member is working in the area of computer science and technologies since 1992.

Six years of development of previous product – the Associative Search Engine (ASM) – linguistic text search system for web data with closed cycle including crawling, indexation, search and management of a millions documents per installation as high specialize service applications written on C++.

More than three years of development of the HCE project products and technologies.