About Nutch

Table of contents

1. Overview

Nutch is open source web-search software. It builds on Lucene Java, adding web-specifics, such as a crawler, a link-graph database, parsers for HTML and other document formats, etc.

For more information about Nutch, please see the Nutch wiki.