test

	<!DOCTYPE html>
	<html lang="en">
	<head>
	<title>LangChain <> Unstructured</title>
	<meta charset="utf-8" />
	<meta http-equiv="X-UA-Compatible" content="IE=edge" />
	<meta name="HandheldFriendly" content="True" />
	<meta name="viewport" content="width=device-width, initial-scale=1.0" />
	<link rel="stylesheet" type="text/css" href="/assets/built/screen.css?v=2df407f2e9" />

	<link rel="icon" href="https://blog.langchain.dev/content/images/size/w256h256/2023/01/parroticon-1.png" type="image/png" />
	<link rel="canonical" href="https://blog.langchain.dev/langchain-unstructured/" />
	<meta name="referrer" content="no-referrer-when-downgrade" />
	<meta property="og:site_name" content="LangChain" />
	<meta property="og:type" content="article" />
	<meta property="og:title" content="LangChain <> Unstructured" />
	<meta property="og:description" content="One of the core value props of LangChain is the ability to combine Large Language Models with your own text data. There are multiple (four!) different methods of doing so, and many different applications this can power.
	A step that sits upstream of using text data is the ability to" />

	<meta property="og:url" content="https://blog.langchain.dev/langchain-unstructured/" />
	<meta property="og:image" content="https://blog.langchain.dev/content/images/2023/02/Screen-Shot-2023-02-05-at-11.32.11-PM.png" />
	<meta property="article:published_time" content="2023-02-06T07:32:46.000Z" />
	<meta property="article:modified_time" content="2023-06-24T22:48:36.000Z" />
	<meta name="twitter:card" content="summary_large_image" />
	<meta name="twitter:title" content="LangChain <> Unstructured" />
	<meta name="twitter:description" content="One of the core value props of LangChain is the ability to combine Large Language Models with your own text data. There are multiple (four!) different methods of doing so, and many different applications this can power.

	A step that sits upstream of using text data is the ability to" />
	<meta name="twitter:url" content="https://blog.langchain.dev/langchain-unstructured/" />
	<meta name="twitter:image" content="https://blog.langchain.dev/content/images/2023/02/Screen-Shot-2023-02-05-at-11.32.11-PM.png" />
	<meta name="twitter:label1" content="Written by" />
	<meta name="twitter:data1" content="LangChain" />
	<meta name="twitter:site" content="@LangChainAI" />
	<meta name="twitter:creator" content="@LangChainAI" />
	<meta property="og:image:width" content="1335" />
	<meta property="og:image:height" content="710" />
	<script type="application/ld+json">
	{
	"@context": "https://schema.org",

	"@type": "Article",
	"publisher": {
	"@type": "Organization",
	"name": "LangChain",
	"url": "https://blog.langchain.dev/",
	"logo": {
	"@type": "ImageObject",
	"url": "https://blog.langchain.dev/content/images/size/w256h256/2023/01/parroticon-1.png",
	"width": 60,
	"height": 60
	}
	},
	"author": {
	"@type": "Person",
	"name": "LangChain",
	"image": {
	"@type": "ImageObject",
	"url": "https://blog.langchain.dev/content/images/2023/01/parroticon.png",

	"width": 448,
	"height": 448
	},
	"url": "https://blog.langchain.dev/author/langchain/",
	"sameAs": [
	"https://twitter.com/LangChainAI"
	]
	},
	"headline": "LangChain <> Unstructured",
	"url": "https://blog.langchain.dev/langchain-unstructured/",
	"datePublished": "2023-02-06T07:32:46.000Z",
	"dateModified": "2023-06-24T22:48:36.000Z",
	"image": {
	"@type": "ImageObject",

	"url": "https://blog.langchain.dev/content/images/2023/02/Screen-Shot-2023-02-05-at-11.32.11-PM.png",
	"width": 1335,
	"height": 710
	},
	"description": "One of the core value props of LangChain is the ability to combine Large Language Models with your own text data. There are multiple (four!) different methods of doing so, and many different applications this can power.\n\nA step that sits upstream of using text data is the ability to get your data into a text form. This can be rather tricky due to the multitude of different formats that exist out there.\n\nEnter... unstructured.io.\n\nUnstructured is a company with a mission of transforming natural l",
	"mainEntityOfPage": {
	"@type": "WebPage",
	"@id": "https://blog.langchain.dev/"
	}
	}
	</script>

	<meta name="generator" content="Ghost 5.30" />
	<link rel="alternate" type="application/rss+xml" title="LangChain" href="https://blog.langchain.dev/rss/" />
	<script defer src="https://cdn.jsdelivr.net/ghost/sodo-search@~1.1/umd/sodo-search.min.js" data-key="eaa48df0e56a1183c31a3a39e8" data-styles="https://cdn.jsdelivr.net/ghost/sodo-search@~1.1/umd/main.css" data-sodo-search="https://blog.langchain.dev/" crossorigin="anonymous"></script>
	<link href="https://blog.langchain.dev/webmention" rel="webmention" />
	<script defer src="/public/cards.min.js?v=2df407f2e9"></script><style>:root {--ghost-accent-color: #12133a;}</style>
	<link rel="stylesheet" type="text/css" href="/public/cards.min.css?v=2df407f2e9">

	</head>
	<body class="post-template is-head-left-logo has-cover">
	<header id="gh-head" class="gh-head outer">
	
	LangChain
	
	<button class="gh-search gh-icon-btn" data-ghost-search><svg xmlns="http://www.w3.org/2000/svg" fill="none" viewBox="0 0 24 24" stroke="currentColor" stroke-width="2" width="20" height="20"><path stroke-linecap="round" stroke-linejoin="round" d="M21 21l-6-6m2-5a7 7 0 11-14 0 7 7 0 0114 0z"></path></svg></button>
	<button class="gh-burger"></button>

	<nav class="gh-head-menu">
	<ul class="nav">
	<li class="nav-home">Home
	<li class="nav-about">About
	<li class="nav-github">GitHub
	<li class="nav-docs">Docs
	</nav>

	<button class="gh-search gh-icon-btn" data-ghost-search><svg xmlns="http://www.w3.org/2000/svg" fill="none" viewBox="0 0 24 24" stroke="currentColor" stroke-width="2" width="20" height="20"><path stroke-linecap="round" stroke-linejoin="round" d="M21 21l-6-6m2-5a7 7 0 11-14 0 7 7 0 0114 0z"></path></svg></button>
	</header>
	<main id="site-main" class="site-main">
	<article class="article post ">
	<header class="article-header gh-canvas">

	<h1 class="article-title">LangChain <> Unstructured</h1>
	<section class="article-byline-content">
	<ul class="author-list">
	<li class="author-list-item">
	
	
	

	<h4 class="author-name">LangChain</h4>
	<time class="byline-meta-date" datetime="2023-02-05">Feb 5, 2023</time>
	• 1 min read
	</section>
	<figure class="article-image">
	<img
	srcset="/content/images/size/w300/2023/02/Screen-Shot-2023-02-05-at-11.32.11-PM.png 300w,
	/content/images/size/w600/2023/02/Screen-Shot-2023-02-05-at-11.32.11-PM.png 600w,

	/content/images/size/w1000/2023/02/Screen-Shot-2023-02-05-at-11.32.11-PM.png 1000w,
	/content/images/size/w2000/2023/02/Screen-Shot-2023-02-05-at-11.32.11-PM.png 2000w"
	sizes="(min-width: 1400px) 1400px, 92vw"
	src="/content/images/size/w2000/2023/02/Screen-Shot-2023-02-05-at-11.32.11-PM.png"
	alt="LangChain <> Unstructured"
	/>
	</figure>
	</header>
	<section class="gh-content gh-canvas">
	<p>One of the core value props of LangChain is the ability to combine Large Language Models with your own text data. There are multiple (four!) different methods of doing so, and <a

href="https://langchain.readthedocs.io/en/latest/use_cases/question_answering.html">many different applications this can power. A step that sits upstream of using text data is the ability to get your data into a text form. This can be rather tricky due to the multitude of different formats that exist out there. Enter... unstructured.io. Unstructured is a company with a mission of transforming natural language data from raw to machine ready. One of the main ways they do this is with an open source Python package. This package as support for MANY different types of file extensions: <code>.txt</code>, <code>.docx</code>, <code>.pptx</code>, <code>.jpg</code>, <code>.png</code>, <code>.eml</code>, <code>.html</code>, and <code>.pdf</code> documents. After playing around with

Unstructured, we realized that by integrating with it we could easily start to build out first class support for loading documents of all types into a format that LangChains could work with. So we created the Document Loaders module, a large part of which is powered by Unstructured. There are currently two loaders that are powered by Unstructured. Both seem rather simple, but are quite powerful. The first is the UnstructuredFileLoader. This has a simple interface (you just pass it a file path) but under the hood Unstructured is doing a lot of smart logic to infer which data type it is (PDF, PowerPoint, image, etc) and extract text. The second is the Directory Loader. Again, this has a pretty simple interface: it takes only a path to a directory and an optional regex to glob for files against. But under the hood it is looping over all files and using the above UnstructuredFileLoader to

	load them. This makes it possible to load files of all types in a single call. We're incredibly excited to have made this integration with Unstructured. With their focus on transforming raw data into clean text, it makes it incredibly easy to combine language models with your data, no matter what form it is in.
	</section>
	</article>
	</main>
	<aside class="read-more-wrap outer">
	<article class="post-card post">

	
	<img class="post-card-image"
	srcset="/content/images/size/w300/2023/06/Langchain--1-.jpg 300w,
	/content/images/size/w600/2023/06/Langchain--1-.jpg 600w,
	/content/images/size/w1000/2023/06/Langchain--1-.jpg 1000w,
	/content/images/size/w2000/2023/06/Langchain--1-.jpg 2000w"
	sizes="(max-width: 1000px) 400px, 800px"
	src="/content/images/size/w600/2023/06/Langchain--1-.jpg"
	alt="🎉 Prem Challenge🎉"
	loading="lazy"
	/>
	

	
	<header class="post-card-header">
	<h2 class="post-card-title">
	🎉 Prem Challenge🎉
	</h2>
	</header>
	We're excited to announce a challenge hosted by Prem in collaboration with LangChain.
	At LangChain we try to make it easy as possible to experiment with as many different models as possible. That includes the incredible number of emerging open source models. We've made efforts to make our framework as
	

	<footer class="post-card-meta">
	<time class="post-card-meta-date" datetime="2023-06-26">Jun 26, 2023</time>
	6 min read
	</footer>
	</article>
	<article class="post-card post">
	
	<img class="post-card-image"
	srcset="/content/images/size/w300/2023/06/mongo.webp 300w,
	/content/images/size/w600/2023/06/mongo.webp 600w,
	/content/images/size/w1000/2023/06/mongo.webp 1000w,

	/content/images/size/w2000/2023/06/mongo.webp 2000w"
	sizes="(max-width: 1000px) 400px, 800px"
	src="/content/images/size/w600/2023/06/mongo.webp"
	alt="LangChain <> MongoDB Atlas"
	loading="lazy"
	/>
	
	
	<header class="post-card-header">
	<h2 class="post-card-title">
	LangChain <> MongoDB Atlas

	</h2>
	</header>
	Today we're announcing LangChain's integration with MongoDB Atlas, adding support for one of the most popular developer data platforms in the world. This is an integration so anticipated that a few developers added the integration before we were ready to announce it :)
	Overview
	One of the key components
	
	<footer class="post-card-meta">
	<time class="post-card-meta-date" datetime="2023-06-22">Jun 22, 2023</time>
	3 min read
	</footer>

	</article>
	<article class="post-card post">
	
	<img class="post-card-image"
	srcset="/content/images/size/w300/2023/06/teaser_chatbot.jpg 300w,
	/content/images/size/w600/2023/06/teaser_chatbot.jpg 600w,
	/content/images/size/w1000/2023/06/teaser_chatbot.jpg 1000w,
	/content/images/size/w2000/2023/06/teaser_chatbot.jpg 2000w"
	sizes="(max-width: 1000px) 400px, 800px"
	src="/content/images/size/w600/2023/06/teaser_chatbot.jpg"
	alt="Data-Driven Characters"

	loading="lazy"
	/>
	
	
	<header class="post-card-header">
	<h2 class="post-card-title">
	Data-Driven Characters
	</h2>
	</header>
	Data-driven-characters is a repo for creating, debugging, and interacting your own chatbots conditioned on your own story corpora.
	

	<footer class="post-card-meta">
	<time class="post-card-meta-date" datetime="2023-06-19">Jun 19, 2023</time>
	11 min read
	</footer>
	</article>
	</aside>
	<footer class="site-footer outer">

	<section class="copyright">LangChain © 2023</section>
	<nav class="site-footer-nav">
	<ul class="nav">
	<li class="nav-sign-up">Sign up
	</nav>
	Powered by Ghost
	</footer>
	<script
	src="https://code.jquery.com/jquery-3.5.1.min.js"

	integrity="sha256-9/aliU8dGd2tb6OSsuzixeV4y/faTqgFtohetphbbj0="
	crossorigin="anonymous">
	</script>
	<script src="/assets/built/casper.js?v=2df407f2e9"></script>
	<script>
	$(document).ready(function () {
	// Mobile Menu Trigger
	$('.gh-burger').click(function () {
	$('body').toggleClass('gh-head-open');
	});
	// FitVids - Makes video embeds responsive
	$(".gh-content").fitVids();
	});
	</script>
	</body>

	</html>