November 2025

Pay-To-Crawl

A Creative Commons Issue Brief: Backgrounders on topics related to AI & the commons

Pay-to-crawl refers to emerging technical systems used by websites to automate compensation for when their digital content—such as text, images, and structured data—is accessed by machines.

Introduction

While machine access of digital content is not entirely new, pay-to-crawl systems have emerged in response to the disruption caused by large artificial intelligence (AI) models accessing vast amounts of content without permission, attribution, or compensation.

Pay-to-crawl systems are described as addressing issues including increased hosting costs for websites, reductions in traffic and visibility brought about by AI-enabled search, and the undermining of referral and advertising-based business models.

Viewed through a wider lens, pay-to-crawl systems represent one of the latest incarnations of content monetization and control, combining elements of paywall, digital rights management (DRM), and micropayment approaches.

How Pay-To-Crawl Systems Work

Not all pay-to-crawl systems work in the same way. Some systems, for example, are focused more on blocking machines from accessing content than making them pay. However, most systems tend to involve some combination of the following components:

- **Authentication**: Pay-to-crawl systems require the person, organization, or product operating a machine seeking to access content to identify themselves. Many systems use cryptographic authentication rather than methods that operators of machines have proven able to circumvent in the past, such as user agent strings or IP addresses.

- **Access Control**: Pay-to-crawl systems set granular and functional rules to define which machines can access content, under what conditions, and whether access is freely enabled, blocked, or billed. Some systems allow websites to set rate limits, rather than fully block off access.

- **Pricing & Contracting**: Pay-to-crawl systems define compensation for access, such as per page or by data volume, or on a subscription basis. Contracts are generally automated, sometimes using standardized licenses or terms. Terms are not only or always financial—they can involve attribution and other reuse obligations. Some systems enable collective bargaining on behalf of groups of websites.

- **Payment**: Pay-to-crawl systems provide mechanisms for payment, often using secure third-party processing services. Payment can be made directly and immediately to the website, or taken by an operator of the system, such as a web services provider, on their behalf.

- **Content Delivery**: Upon authentication and payment, pay-to-crawl systems enable access to content, typically in formats optimized for machine consumption. Some systems enable encrypted access to non-public content.

- **Metering & Logging**: Pay-to-crawl systems often log information related to access and use of content to enable billing and some degree of auditability.

The role of websites in developing and using pay-to-crawl systems varies. Some may choose to deploy such a system themselves, using emerging protocols and code. In other cases, pay-to-crawl systems are being developed as specialist, paid-for products, or introduced by web services providers (such as domain hosts and content delivery networks) on behalf of websites. As a result, pay-to-access systems vary in terms of their openness, standardization, and interoperability, as well as the permissiveness of their access controls and payment terms.

Considerations

In the face of unprecedented consumption of digital content by large AI models—both in scale and in impact—the use of pay-to-crawl systems may help websites sustain the creation and publication of content, or tackle what they consider to be substitutive uses of their works. However, overbroad and indiscriminate use of pay-to-crawl systems could block off access to digital

content for researchers, nonprofits, cultural heritage institutions, educators, and other actors working in the public interest; obstruct legitimate uses of content protected by copyright or other laws; and create new walled gardens, web gatekeepers, and excesses of power. Wide adoption of pay-to-crawl could ultimately represent a shift away from the spirit of the open web towards a more tightly controlled and monetized content ecosystem.

Examples

Examples of pay-to-crawl systems and related initiatives include Pay Per Crawl by Cloudflare, AI RevShare by Valyu, GistAttribution by ProRata, Open Licensing Protocol by RSL and TollBit.

Notes on Terminology

We're choosing to use *pay-to-crawl* to describe these systems on account of the term already being widely used. We generally prefer the broader term *pay-to-access*, given that, technically speaking, there are many purposes and forms of machine access to content beyond crawling. *Crawling* does not, for example, adequately describe the process of extracting and making copies of content (often referred to as *scraping*), nor analyzing them to derive insights or patterns (*text and data mining*).

Websites is a broad category. The term *publisher* might be more appropriate to describe the entity responsible for the content and the user of a pay-to-crawl system, especially in domains such as news, academia, and the media. The user of a pay-to-crawl system, regardless of how they are described, is not always the original creator or owner of the content such a system is used to manage access to.

In this context, *machines* refers to the systematic access and use of digital content using code and automated programs, rather than typical human browsing and consumption. *Bots* is also sometimes used. This shouldn't obscure the fact that code and automated programs are ultimately operated by humans.