Crawl and index

You are going to try and act as a web crawler to determine what would be recorded in a search engine index after the page has been visited.

The pages are presented as HTML source code and have been simplified to let you focus on the key information.

Web page 1: Wikipedia page about Scratch

(Creative Commons Attribution-ShareAlike 3.0 Unported License - https://en.wikipedia.org/wiki/Scratch_(programming_language))

```
<head>
<title>Scratch (programming language) - Wikipedia</title>
<meta name="description" content="Information about Scratch, a visual
programming language">
</head>

<body>
<h1>Scratch (programming language)</h1>
<p>Scratch is a block-based visual programming language and website targeted
primarily at children. Users of the site can create online projects using a block-like
interface. The service is developed by the MIT Media Lab, has been translated into 70+
languages, and is used in most parts of the world. Scratch is taught and used in
after-school centers, schools, and colleges, as well as other public knowledge institutions.
As of May 2019, community statistics on the language's official website show more than
40 million projects shared by over 40 million users, and almost 40 million monthly
website visits.</p>
```

```html
<p>Scratch takes its name from a technique used by disk jockeys called "scratching", where vinyl records are clipped together and manipulated on a turntable to produce different sound effects and music. Like scratching, the website lets users mix together different media (including graphics, sound, and other programs) in creative ways by "remixing" projects.</p>
<em>Last update: 20/04/20</em>
</body>
```

Results stored in search engine index

Read the content of the web page as if you were a crawler. Fill in the table below to show what information you believe is most important when cataloguing this page:

Five keywords for page (in order of relevance)	Type of content found (images, text, etc.)	Date of last update
1. 2. 3. 4. 5.		

Web page 2: Online eBook

(R. Adam Dastrup - Creative Commons Attribution 4.0 International License - https://slcc.pressbooks.pub/physicalgeography/chapter/4-1/)

```html
<head>
<title>Chapter 4: Plate Tectonics</title>
<meta name="description"  content="How plate tectonics shaped the world from its beginnings as Pangea">
</head>

<body>
<h1>Early Evidence for Continental Drift</h1>
```

```html
<p>The first piece of evidence is that the shape of the coastlines of some continents fit together like pieces of a jigsaw puzzle. Since the first world map, people have noticed the similarities in the coastlines of South America and Africa, and the continents being ripped apart had even been mentioned as an explanation. Antonio Snider-Pellegrini even did preliminary work on continental separation and matching fossils in 1858.</p>

<p>What Wegener did differently than others was synthesize a significant amount of data in one place, as well as use the shape of the continental shelf, the actual edge of the continent, instead of the current coastline, which fit even better than previous efforts.
</p>
<img src="PMap.jpg"  alt="Map of Pangea">
<em>Last update: May 1, 2019</em>
</body>
```

Results stored in search engine index

Read the content of the web page as if you were a crawler. Fill in the table below to show what information you believe is most important when cataloguing this page:

Five keywords for page (in order of relevance)	Type of content found (images, text, etc.)	Date of last update
1. 2. 3. 4. 5.		