# 🕷️ Crawlability Report – Etsy.com

Project: Intelligent Web Crawler & Analyzer

Team Role: Member 1 – Crawlability Specialist

Assigned Website: https://www.etsy.com

## 🔍 Objective

To analyze Etsy’s robots.txt file, determine crawl permissions, and summarize the rules that govern automated web crawlers.

## 📄 Robots.txt Overview

The robots.txt file for Etsy is available at:

https://www.etsy.com/robots.txt

## ✅ Allowed and ❌ Disallowed Paths

Disallowed Paths (Partial List):

* Disallow: /c/
* Disallow: /listing/
* Disallow: /search/
* Disallow: /your/
* Disallow: /cart/
* Disallow: /checkout/
* Disallow: /people/
* Disallow: /favourite/
* Disallow: /register/
* Disallow: /account/
* Disallow: /email/
* Disallow: /inbox/
* Disallow: /int/
* Disallow: /translations/

These disallowed rules mostly block:

- User-specific sections (/your/, /account/, /cart/)

- Listings and product pages (/listing/)

- Search results (/search/)

- Registration and messaging

## ✅ Allowed Paths (Partial List)

* Allow: /mailinglist/email/
* Allow: /uk/mailinglist/email/
* Allow: /au/mailinglist/email/
* Allow: /ca/mailinglist/email/
* Allow: /de-en/mailinglist/email/
* Allow: /dk-en/mailinglist/email/
* Allow: /fi-en/mailinglist/email/
* Allow: /hk-en/mailinglist/email/
* Allow: /api/v3/ajax/bespoke/public/neu/specs/

## 🔁 Crawl Delay

Not specified. No Crawl-delay directive was found. Crawlers must respect server load responsibly.

## 🔗 Sitemap Links

* https://www.etsy.com/sitemaps.xml
* https://www.etsy.com/blog/sitemap.xml.gz
* https://www.etsy.com/sitemaps/shops/vip-seller.xml
* https://www.etsy.com/fi-en/sitemaps/shops/vip-seller.xml
* https://www.etsy.com/au/sitemaps/shops/vip-seller.xml

## 🧠 Crawling Logic Code Snippet

import urllib.robotparser  
  
rp = urllib.robotparser.RobotFileParser()  
rp.set\_url("https://www.etsy.com/robots.txt")  
rp.read()  
  
test\_url = "https://www.etsy.com/shop/some-shop"  
if rp.can\_fetch("\*", test\_url):  
 print("✅ Allowed to crawl:", test\_url)  
else:  
 print("❌ Disallowed to crawl:", test\_url)

## 🧾 Summary

|  |  |
| --- | --- |
| Feature | Status |
| Crawl Allowed? | Partially (many paths blocked) |
| Crawl Delay | Not Specified |
| Sitemap Provided | ✅ Yes |
| Disallowed Focus | Listings, Search, User Data |
| Allowed Focus | Homepage, Some Categories |