Model Inference for Web Applications

Essentially, all models are wrong, but some are useful.

The Problem

Model Inference

Automatically **abstract** a model (finite state machine) of a Web App

Why?

Model-based Approaches

With a good model in place:

- Testers can easily design test cases;
- Automatically generate test artifacts (e.g.: page objects);
- Automatically generate test cases;
- Apply Formal Verification techniques;

Crawling-based Approaches

Web Application Crawlers

- Software agents that automatically explore a Web Application
- Advanced crawlers also account for async client-side JavaScript
- Build a **Reachability Graph** of the Web App
- The Reachability Graph is the **model**!
- Every reachable page is a **state**

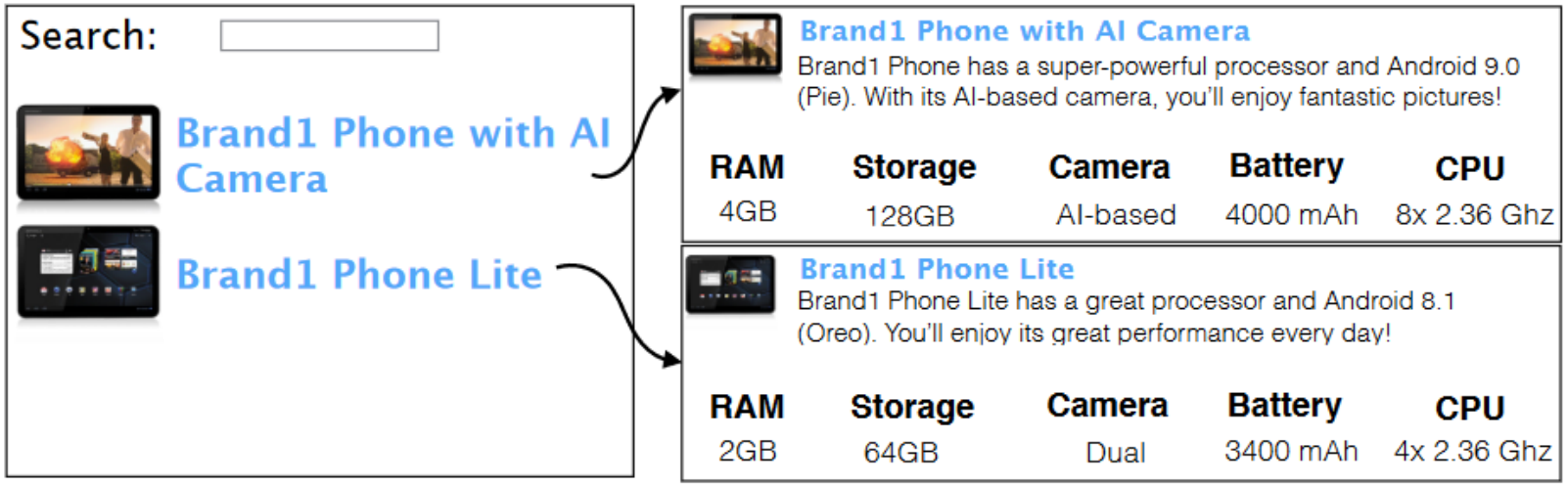
Image via searchengineland.com

The Near-duplicate Problem

Near-duplicate web pages

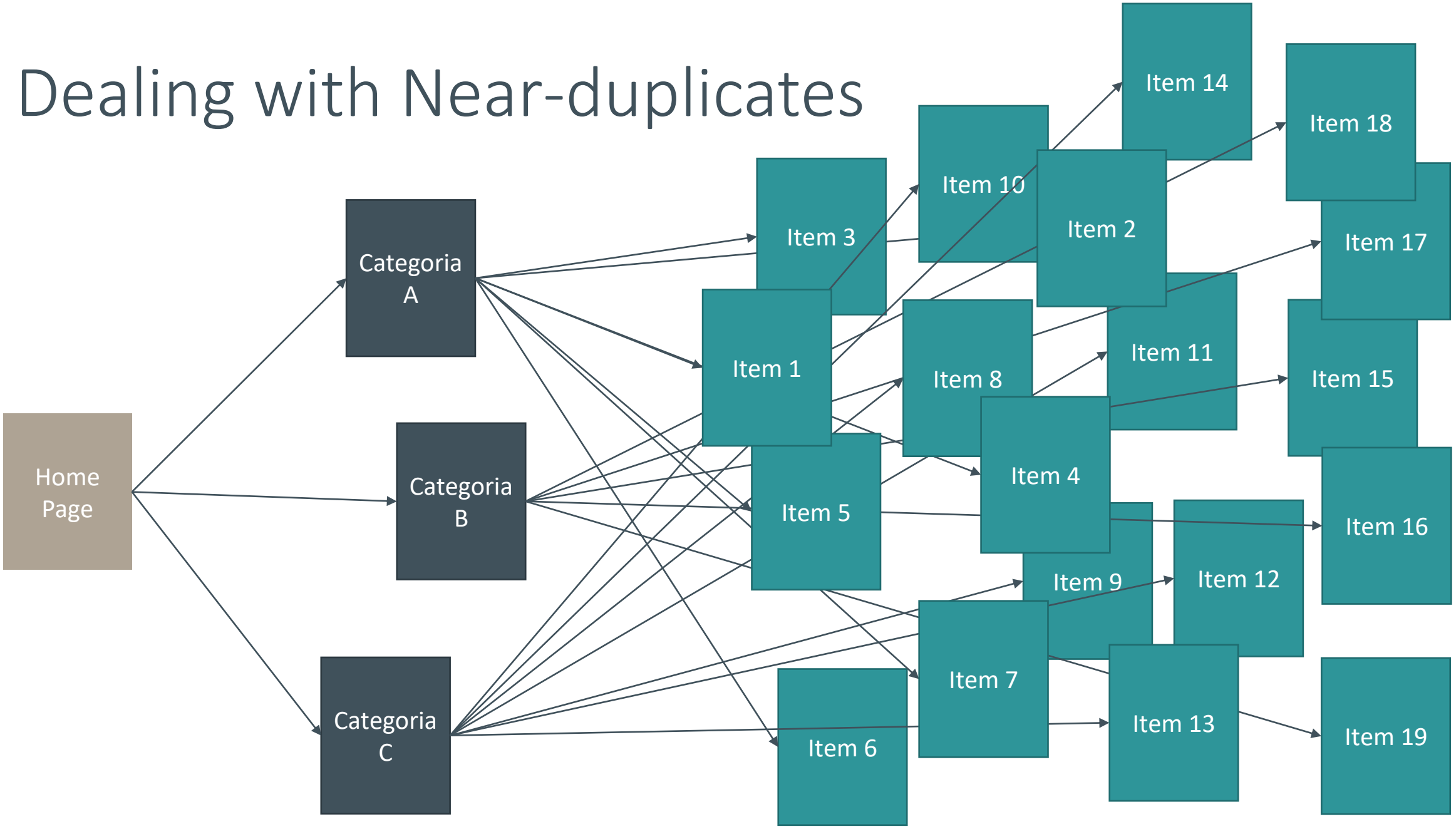
Models abstracted by traditional crawling are **not very good** models

- lots of **near-duplicate** states!

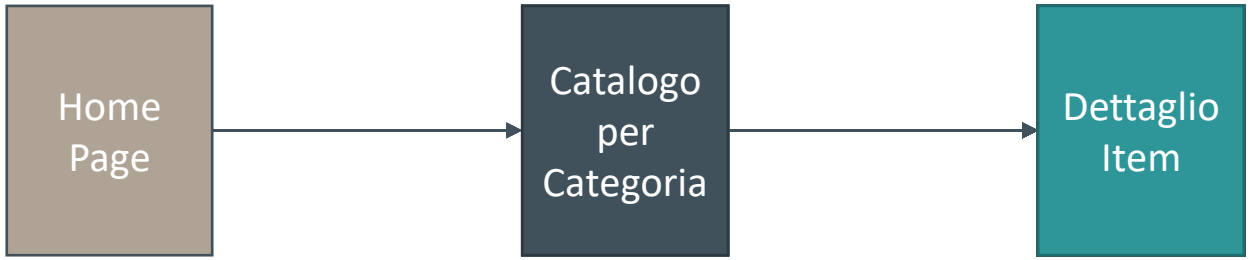


Example of near-duplicate web pages, from Yandrapally et Al. "Near-duplicate detection in web app model inference." *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*. 2020.

Dealing with Near-duplicates



Dealing with Near-duplicates



Detecting Near-duplicates

- How do we detect near-duplicates?
- Not an easy task!
- There are different kinds of near-duplicates

Cyberpunk 2077

59,99 €

BUY NOW

Explore Night City, where the choices you make shape the story and the world around you.

The Witcher 3

14,98 €

BUY NOW

As war rages on throughout the Northern Realms, you take on the greatest contract of ...

Detecting Near-duplicates

- How do we detect near-duplicates?
- Not an easy task!
- There are different kinds of near-duplicates

Category: Horror	
SOMA	4,99 €
Resident Evil 7	29,99 €
Outlast	16,79 €

Category: RPG	
Cyberpunk 2077	59,99 €
RuneScape	Free to play
Baldur's Gate 3	59,99 €
Fallout 4	29,99 €
Torchlight III	33,99 €
Dark Souls III	59,99 €