

Виявлення 4XX та 5XX помилок на сайті як впливний фактор на результат точності рекомендаційних систем

Вінницький національний технічний університет

Анотація

Дано опис рекомендаційних систем на основних впливних факторів на їх точність. Представлено аналіз виявлення 40X та 50X помилок сайту. Поданий огляд використаних інструментів, їх недоліки та переваги. Описаний вплив 40X та 50X помилок на точність рекомендаційних систем на сайт в цілому. Описані методи автоматизації виявлення даного виду помилок.

Ключові слова: рекомендаційна система, точність, 4XX, 5XX.

Abstract

A description of the advisory systems on the main influential factors is given on their accuracy. The analysis of detection of 4XX and 5XX errors of a site is presented. An overview of the tools used, their drawbacks and advantages. The influence of 4XX and 5XX errors on the accuracy of advisory systems on the site as a whole is described. The methods of automation of detection of this kind of errors are described.

Keywords: recommendation system, accuracy, 40X, 50X.

Рекомендаційна система — підклас системи фільтрації інформації, яка будує рейтинговий перелік об'єктів (фільми, музика, книги, новини, веб-сайти), яким користувач може надати перевагу. Для цього використовується інформація з профілю користувача. Існують дві основні стратегії створення рекомендаційних систем: фільтрація вмісту і колаборативна фільтрація. При фільтрації вмісту створюються профілі користувачів і об'єктів. Профілі користувачів можуть містити демографічну інформацію або відповіді на певний набір питань. Профілі об'єктів можуть містити назви жанрів, імена акторів, імена виконавців, тощо. Або якусь іншу інформацію в залежності від типу об'єкта. При колаборативній фільтрації використовується інформація про поведінку користувачів у минулому — наприклад, інформація про придбання або оцінки. В цьому разі не має значення, з якими типами об'єктів ведеться робота, але при цьому можна брати до уваги неявні характеристики, які складно було б врахувати при створенні профілю. Основна проблема цього типу рекомендаційних систем — «холодний старт»: відсутність даних про користувачів чи об'єкти, які нещодавно з'явилися у системі.

Таким чином можна зробити висновок, що продукт як сайт повинен бути максимально корисним для користувача. Цим обґрунтовується актуальність проведення даних досліджень.

На точність рекомендаційних систем впливають як зовнішні фактори так і внутрішні. Внутрішні фактори залежать переважно від кількості вхідних даних, їх точності, методу фільтрації та прогнозування і т.д. Проте, базовими впливними факторами є зовнішні, без оптимізації яких

неможливо добитись результати. До основних зовнішніх впливних факторів точності рекомендаційних систем відносять:

- 4XX та 5XX сторінки (Наприклад 404, 502)
- Швидкість сайту
- Дизайн (юзабіліті)
- Розміщення рекомендаційного блоку
- Наявність редіректів
- Кількість реклами
- Кількість контенту
- Кількість внутрішніх посилань на сайті

Розберемо детальніше кожний фактор. 40X та 50X сторінки (Наприклад 404, 502) є популярною проблемою, якщо наприклад сторінка не знайдена, коли не оптимізована під відсутність товару, якщо не враховувати різні технічні помилки, або 50X помилки сервера, коли немає доступу до сторінки, або до сайту в цілому. В цій тезі доповіді буде висвітлена найбільше сама ця проблема, проте розберемо поверхнево інші фактори.

Швидкість сайту: за дослідженнями компанії Google, майже половина користувачів на мобільній версії сайту покидає сайт, якщо він завантажується довше 3 секунд. Це означає, що який би не був точний та складний алгоритм рекомендаційної системи – половина користувачів на мобільній версії його не побачить. Тому оптимізації швидкості сайту перед розробкою рекомендаційної системи досить важлива проблема, яка є трендом. Оскільки в червні минулого року швидкість сайту вперше стала офіційним фактором ранжування в рамках оновлення Google Speed Update.

Дизайн (юзабіліті) є досить важливим впливним фактором, оскільки кожному користувачеві повинен бути зрозумілим дизайн та навігація сайту. Без цього, людина або одразу покине сайт, або не буде ходити довго по сайту і його сторінках, і таким чином може не звернути увагу на рекомендаційну систему, яка може бути оформлена у форматі блоку, або слайдера.

Розміщення рекомендаційного блоку є теж одним із основних впливних зовнішніх факторів, оскільки часто буває ситуація, коли рекомендаційний блок знаходить внизу сторінки і велика частина користувачів не доскролить до блоку, таким чином його просто не побачить. З цього випливає, що знову, який би не був точний алгоритм рекомендаційної системи – якщо рекомендаційний блок не побачить користувач, то це суттєво погіршує базові результати вимірювання, які зазвичай дорівнюють відношенню кількості користувачів, які зайшли на сторінку, до користувачів які клікнули на блок. Тому, якщо рекомендаційного блоку немає у видимій частині екрану – потрібно аналізувати точність такого блоку як відношення користувачів, які побачили рекомендаційний блок до користувачів, які на нього клікнули.

Наявність редіректів є важливим фактором. Уявімо ситуацію, що рекомендаційний блок закріплений у видимій частині екрану, дизайн та навігація зручна та зрозуміла користувачам, але в кінцевому результаті користувача просто перекидає 301 редіректом на іншу сторінку. Таким чином, потрібно уникати 301 (або 302) редіректів на інші сторінки сайту. Дозволяється тільки робити редірект URL в URL, наприклад із HTTP версії сторінки на HTTPS.

Кількість реклами та контенту теж є впливним фактором на точність. Реклама, якщо вона тареготована може привернути увагу користувача та він може перейти на рекламний сайт. Велика кількість контенту може вплинути на конверсії, оскільки в користувача багато інформації/фото/відео, які можуть спричинити меншу клікабельність по рекомендаційному блоку.

Кількість внутрішніх посилань (перелінковка) є важливим фактором, оскільки чим більше ми даємо користувачу можливостей клікнути десь на сторінці, тим більше вірогідність того, що він цим скористається, якщо посилання на інші сторінки сайту релевантні та присутні тільки в рамках навігації та зручності.

Розглянемо детальніше проблему відповідей сервера:

1. **1xx** - інформаційні коди. Вони відповідають за процес передачі даних. Це тимчасові коди, вони інформують про те, що запит прийнятий і обробка триватиме.
2. **2xx** - успішна обробка. Запит був отриманий і успішно оброблений сервером.
3. **3xx** - перенаправлення (редирект). Ці відповіді сервера свідчать, що потрібно зробити подальші дії для виконання запиту. Наприклад, зробити запит за іншою адресою.
4. **4xx** - помилка користувача. Це означає, що запит не може бути виконаний з його вини.

5. **5xx** - помилка сервера. Ці коди виникають через помилки на стороні сервера. В даному випадку користувач все зробив правильно, але сервер не може виконати запит. Для кодів цього класу сервер обов'язково показує повідомлення, що не зрозумів запит і з якої причини.

Найпопулярніший і важливий статус. Чи означає, що запит виконаний успішно відповідно до очікувань користувача - запитані дані або сторінка існують і доступні для перегляду. Всі сторінки, які ми хочемо бачити в індексі пошукових систем, повинні віддавати код 200 ОК. Стандартним методом можна перевірити відповідь сервера:

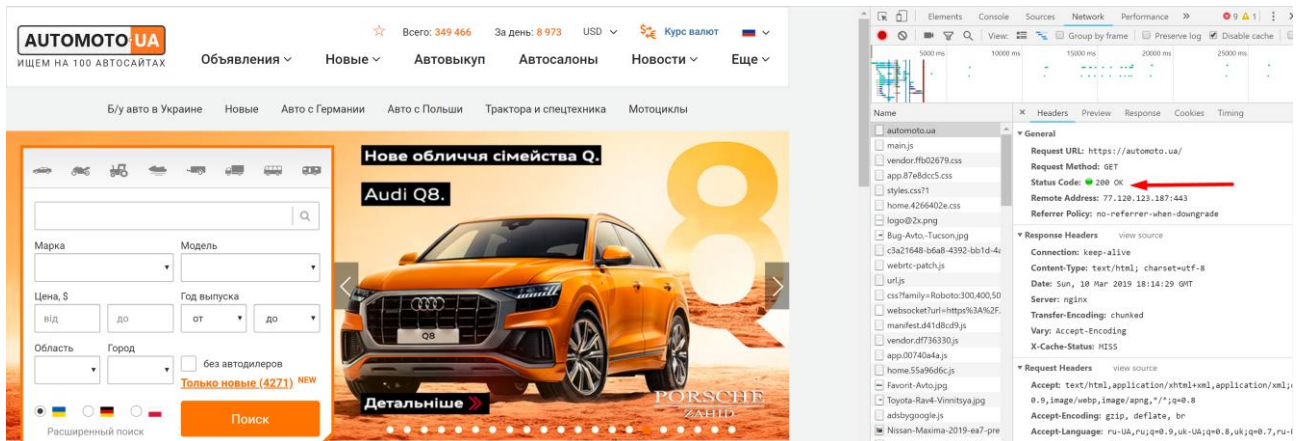


Рисунок 1 – відповідь сервера 200 ОК

Відповідь 301 – це редірект, примусове перенаправлення на іншу сторінку. Найпопулярніше застосування з редірект з http на https версію сайту (Рис 2).

ПРОВЕРКА РЕДИРЕКТА

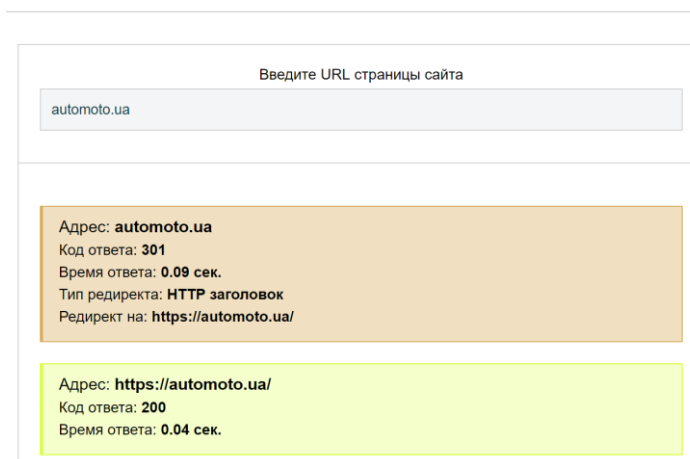


Рисунок 2 – Редірект з HTTP на HTTPS версію сайту

Відповідь 404 – сторінка не знайдена. Таку помилку можна зустріти як через вину сайту (посилання на сайті, яке веде на неіснуючу сторінку), та і через користувача (неправильно написана адреса сайту/сторінки). Виявлення та виправлення з технічної точки зору таких сторінок впливає на точність рекомендаційних систем, оскільки якщо помилково поставити биті (неіснуючі/неправильні) посилання на кінцеві сторінки оголошень, де знаходяться рекомендаційні системи, то користувачі не побачать взагалі рекомендаційні системи, статистики не буде.

Самим популярним методом виявлення таких помилок є інструмент – Google Search Console.

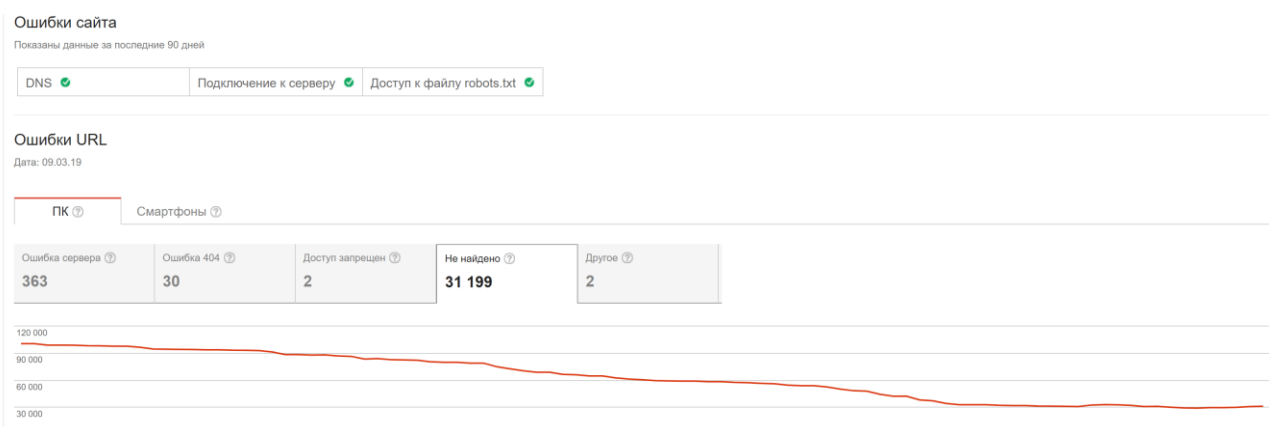


Рисунок 3 – помилки 404 в Google Search Console

504 - Шлюз не відповідає. Відповідь з'являється, якщо сервер працював в якості проксі і не дочекався відповіді від висхідного сервера для завершення запиту. Найбільш критичніша помилка, оскільки переважно вона стосується всього сайту. Таким чином, якщо попавши на 404 сторінку можна знайти іншу, існуючу. То якщо код 504 – сайт може повністю не працювати. Це дуже негативно впливає на користувацький досвід та на ранжування сайту в пошукових системах (SEO)



Рисунок 4 - помилки 504 в Google Search Console

Проте автоматизація виявлення 4XX та 5XX помилок через Google Search Console не можлива. Тому в рамках наукової роботи було використано на налаштовано інструмент Netpeak Spider, який запускає робота на сторінки сайту та аналізує десяток технічних показників сайту (Рис. 5).

1	#	Ссылка	50 Исходящие	Входящие	Статус ссы	Конечная	Код ответа сервера	Content-Тип	Robots.txt	Конечный	Canonical	Refresh	X-Robots-T	Meta Robots
14907	14906	https://img.automoto.ua/Dongfeng-244-undefined-none-	0	0	0	Незв'язанні	200 OK	image/webp	allowed					
14908	14907	https://img.automoto.ua/Dongfeng-244-undefined-none-	0	0	0	Незв'язанні	200 OK	image/webp	allowed					
14909	14908	https://img.automoto.ua/Welger-RP-undefined-none-201	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14910	14909	https://img.automoto.ua/overview/Hyundai-ioniq-2017-af	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14911	14910	https://img.automoto.ua/overview/Renault-Clio-2016-shc	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14912	14911	https://automoto.ua/overview/Hyundai-ioniq-2017-838.ht	0	1	0	Незв'язанні	200 OK & Canonicalized	text/html	ch, allowed			https://auton		
14913	14912	https://automoto.ua/overview/Renault-Clio-2016-843.htn	0	1	0	Незв'язанні	200 OK & Canonicalized	text/html	ch, allowed			https://auton		
14914	14913	https://img.automoto.ua/overview/ravon-Genra-2017-sh	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14915	14914	https://img.automoto.ua/overview/ravon-R2-2017-show-i	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14916	14915	https://automoto.ua/images/nopphoto/1b.jpg	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14917	14916	https://automoto.ua/upload/prototypes/210x0/1728/260-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14918	14917	https://automoto.ua/upload/prototypes/210x0/1730/272-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14919	14918	https://img.automoto.ua/overview/ravon-R4-2017-show-i	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14920	14919	https://automoto.ua/overview/ravon-Genra-2017-842.ht	0	1	0	Незв'язанні	200 OK & Canonicalized	text/html	ch, allowed			https://auton		
14921	14920	https://automoto.ua/overview/ravon-R2-2017-841.html?#	0	1	0	Незв'язанні	200 OK & Canonicalized	text/html	ch, allowed			https://auton		
14922	14921	https://img.automoto.ua/overview/MINI-Clubman-2015-s	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14923	14922	https://automoto.ua/overview/ravon-R4-2017-839.html?#	0	1	0	Незв'язанні	200 OK & Canonicalized	text/html	ch, allowed			https://auton		
14924	14923	https://automoto.ua/overview/MINI-Clubman-2015-840.h	0	1	0	Незв'язанні	200 OK & Canonicalized	text/html	ch, allowed			https://auton		
14925	14924	https://automoto.ua/upload/prototypes/210x0/1720/258-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14926	14925	https://automoto.ua/upload/prototypes/210x0/1639/3-E9	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14927	14926	https://automoto.ua/upload/prototypes/210x0/1633/1-Cc	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14928	14927	https://automoto.ua/upload/prototypes/210x0/1765/850-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14929	14928	https://automoto.ua/upload/prototypes/210x0/1736/418-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14930	14929	https://automoto.ua/upload/prototypes/210x0/1734/391-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14931	14930	https://automoto.ua/upload/prototypes/210x0/1755/698-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14932	14931	https://automoto.ua/upload/prototypes/210x0/1685/81-S	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14933	14932	https://automoto.ua/upload/prototypes/210x0/1692/91-2	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14934	14933	https://automoto.ua/upload/prototypes/210x0/1750/676-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14935	14934	https://automoto.ua/upload/prototypes/210x0/1767/870-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14936	14935	https://automoto.ua/upload/prototypes/210x0/1747/497-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14937	14936	https://automoto.ua/upload/prototypes/210x0/1763/778-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14938	14937	https://automoto.ua/upload/prototypes/210x0/1785/1011	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14939	14938	https://automoto.ua/upload/prototypes/210x0/1706/156-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14940	14939	https://automoto.ua/upload/prototypes/210x0/1784/1000	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14941	14940	https://automoto.ua/upload/prototypes/210x0/1700/103-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14942	14941	https://automoto.ua/upload/prototypes/210x0/1773/903-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14943	14942	https://automoto.ua/upload/prototypes/210x0/1680/45-2	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14944	14943	https://automoto.ua/upload/prototypes/210x0/1758/738-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14945	14944	https://automoto.ua/upload/prototypes/210x0/1681/53-1	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14946	14945	https://automoto.ua/upload/prototypes/210x0/1714/222-	0	0	0	Незв'язанні	404 Not Found	text/html	ch, allowed					
14947	14946	https://automoto.ua/uploads/file/e2/d6/0b/60/alpha.png	0	0	0	Незв'язанні	404 Not Found	image/png	allowed					
14948	14947	https://img.automoto.ua/img/aHR0cHM6Ly9jZGU4cmh	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14949	14948	https://automoto.ua/uploads/file/3d/98/29/2c/AC-10go-19	0	0	0	Незв'язанні	200 OK	image/png	allowed					
14950	14949	https://img.automoto.ua/img/aHR0cHM6Ly9jZGU4cmh	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14951	14950	https://img.automoto.ua/img/aHR0cHM6Ly9jZGU4cmh	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					
14952	14951	https://automoto.ua/uploads/file/2f/39/9d/0a/alpha.romio	0	0	0	Незв'язанні	200 OK	image/png	allowed					
14953	14952	https://img.automoto.ua/img/aHR0cHM6Ly9jZGU4cmh	0	0	0	Незв'язанні	200 OK	image/jpeg	allowed					

Рисунок 5 – Приклад отриманого звіту Netpeak Spider у форматі .xml

Таким чином автоматизуючи процес виявлення 4XX та 5XX помилок, в рамках досліджень ми зможемо виявляти проблемні сторінки та виправляти помилки. Враховуючи, що технічні помилки будуть виправлені, це суттєво покращить точність рекомендаційних систем, оскільки загальна точність рахується як відношення всіх користувачів на сайті до кліку по рекомендаційному блоку.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Рекомендаційна система [Електронний ресурс] // uk.wikipedia.org. – 2016. – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%A0%D0%B5%D0%BA%D0%BE%D0%BC%D0%B5%D0%BD%D0%B4%D0%B0%D1%86%D1%96%D0%B9%D0%BD%D0%B0_%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0.
2. Список кодів стану HTTP [Електронний ресурс] // uk.wikipedia.org. – 2012. – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%A1%D0%BF%D0%B8%D1%81%D0%BE%D0%BA_%D0%BA%D0%BE%D0%B4%D1%96%D0%B2_%D1%81%D1%82%D0%B0%D0%BD%D1%83_HTTP
3. Ответы сервера [Електронний ресурс] // netpeak.net – 2015. – Режим доступу до ресурсу: <https://netpeak.net/ru/blog/otvety-servera-podrobnaya-instruktsiya/>

Кучерук Володимир Юрійович – д.т.н., професор, завідувач кафедри МПА. Місто Вінниця, Вінницький національний технічний університет.

Глушко Михайло Васильович – аспірант, група АС-18, факультет Комп'ютерних систем і автоматики, кафедра МПА. Місто Вінниця, Вінницький національний технічний університет.