

Інтелектуальна інформаційна система автоматичної генерації новин

Василюк Андрій

Кафедра ІСМ
НУ "Львівська політехніка"
Львів, Україна
zoso.vasyluk@gmail.com

Гришканич Роман

Кафедра ІСМ
НУ "Львівська політехніка"
Львів, Україна
uaromix@gmail.com

Intelligent system of automatic news generation able to provide high quality, truthful and actual updates of content for web site, that significantly simplify the work of site editor. Well-designed administration panel allow easy to manage the information system of the people who do not have knowledge in web development. With the obtained results, we can significantly increase traffic to your site, that will give more profits.

Ключові слова: інформаційна система, інтелектуальна система, інтернет новини, генерація новин, шаблон, нейронні мережі, парсинг даних, веб розробка.

ВСТУП

У зв'язку із стрімким розвитком інформаційних технологій, широкої популярності набули Інтернет засоби масової інформації. Люди все більше і більше часу приділяють комп'ютеру ніж телевізору, як це було раніше. Саме цей фактор може дуже допомогти початківцям журналістам, адже вони можуть розпочати свою кар'єру без серйозних капітальних вкладень.

Основним джерелом доходів, у сфері Інтернет ЗМІ, є реклама. А доходи від реклами вже залежить від відвідуваності веб-сайту. На початкових етапах популярність сайт залежить від вдалості SEO, а вже як регулярно заходитимуть до вас користувачі, залежить від актуальності інформації, її достовірності та своєчасності.

Для того щоб забезпечити, оперативне оновлення контенту, редакторам сайту потрібно постійно спостерігати за всім, що твориться у світі, що дуже прив'язує їх до своїх обов'язків. А можна вдатись до хитрощів, тим самим певним чином автоматизувати наповнення, завдяки

використанню систем які будуть наповнювати сайт контентом.

Багато веб-сайтів просто дублюють новини із інших сайтів, але зазвичай користувачі які помічають, що цей сайт дублює контент інших, рідше заходить на цей сайт наступного разу, а тим більше й не додають його у закладки.

Щоб забезпечити унікальність новин потрібно більше ніж просто парсер. Тут потрібно вже використання певної інтелектуальної системи.

ПОСТАНОВКА ПРОБЛЕМИ ТА ШЛЯХИ ЇЇ ВИРІШЕННЯ

Про готові інтелектуальні інформаційні системи автоматичної генерації новин відомо дуже мало, оскільки ті хто їх розробляє та використовує, приховує цей факт. А це є й цілком розумно бо, коли відвідувачі дізнаються, що новини генеруються не людьми, то вони в деякій мірі отримують негативне враження про сайт, а також сайтам конкурентам також не бажано знати про такі розробки, адже це мотивуватиме їх до створення аналогічних систем. У зв'язку з цим полягає проблема у створенні такої інтелектуальної системи.

Для розробки інформаційної системи обрано мову програмування PHP на базі сервера Apache 2 та базу даних MySQL. Це дасть змогу розмістити інформаційну систему із більшості популярних хостинг провайдерів.

Основними частинами інформаційної системи є парсер даних, обробник інформації, шаблонізатор тексту та контролер.

Парсер даних – це частина програми яка відповідає за добування інформації, і яка буде використовуватись для подальшої генерації

новин. Для добування деяких даних можна скористатись API сервісами (здебільшого статистична інформація), проте більшість потрібно інформації знаходиться на звичайних веб сторінках. У таких випадках можна використовувати звичайні бібліотеки для пошуку елементів на веб сторінці, проте суттєвим недоліком цього є те що під кожний сайт прийдеться писати індивідуальний парсер, і тим більше він буде дуже нестійким до оновлення структури сторінки сайту. У даному випадку найкращим рішенням буде підключити нейронні мережі до інформаційної системи, що дасть змогу використовувати універсальний парсер для всіх веб сайтів.

Обробник даних підготовку даних для подальшої шаблонізації. Деякі новини залежать не тільки від щойно добутої даних, а й і від даних які отримали раніше. У таких випадках обробник інформації робить запити до бази даних, де отримує необхідну інформацію, при цьому записує туди поточні дані.

Шаблонізатор даних задає структуру для тексту новин. Для шаблонізації використовуємо TWIG шаблони. Вони є зручні, простими у використанні, а також мають у собі багато можливостей. В панелі адміністрування інформаційної системи додаємо можливість редагування шаблонів. Що дасть можливість редагувати шаблони користувачам які не мають відношення до веб розробки.

Контролер є центральною ланкою усіх частин програми. Завданням контролера є запуснути потрібний парсер у потрібний момент часу, передати парсеру необхідні параметри, отриману з парсера інформацію передати на обробник інформації, дані з обробника передати на шаблонізатор, після чого імпортувати текст новини у базу даних системи керування вмістом веб сайту.

Користувачі в панелі адміністрування обирають додаток парсингу та його режим і налаштовують періодичність чи разове заплановане виконання. Але у багатьох випадках запуск дата і час виконання парсингу виконується із попередньо отриманих результатів парсингу (наприклад потрібно опублікувати результати спортивних подій після їх завершення, то для початку необхідно отримати розкладу спортивних подій

шляхом парсингу веб сайту спортивного турніру). Тож для більшої автоматизації роботи контролер також можна підключити для нейронної мережі, та шляхом попереднього навчання адаптувати його для автоматизованої роботи.

ВИСНОВКИ

Добре проведене навчання для нейронної мережі дає хороший ефект, що дозволяє у значній мірі автоматизувати процеси які мають властивість виконувати завдання в яких немає попередньо передбаченого алгоритму роботи та дозволяють автоматично приймати рішення.

Інтелектуальна система автоматичної генерації новин спроможна забезпечити якісне, своєчасне та актуальне оновлення контенту для веб сайту, а головне, що значно спростить працю редакторів сайту. Добре спроектована панель адміністрування дозволить з легкістю керувати інформаційною системою людьми, які не мають знань у галузі веб розробки.

За допомогою отриманих результатів можна значно підвищити відвідуваність веб сайту, що дозволить збільшити прибутки.

ЛІТЕРАТУРА

- [1] P.J. Hancock, "Data representation in neural nets: an empirical study", Carnegie Mellon University, pp. 36-46, June 2014.
- [2] T.S. Orlando, T.P. Mann, "International Joint Conference on Neural Networks", Washington D.C., IEEE & INNS, pp. 263-266, June 2016.
- [3] D.F. Specht, "Vectorcardiographic diagnosis using the polynomial discriminant method of pattern recognition" Department of Computer Science and Operations Research, North Dakota State University, pp. 12-95, October 2014
- [4] A.A. Waibel, "Consonant recognition by modular construction of large phonetic time-delay neural networks. Neural Information Processing Systems", San Mateo, Kaufmann Publishers, pp. 215-223, Decemder 2013.
- [5] G. B. Josin, "Development of a neural network autopilot model ford high performance aircraft", International Joint Conference on Neural Networks, pp. 15-19, January 2016.