

## Использование машинного обучения для продвижения сайтов.

*И.А. Богомолов, А.С. Дмитриев, Ю. В. Киселев*

*Волгоградский государственный технический университет*

**Аннотация:** Поисковая оптимизация позволяет сайту выходить в поисковых системах на более высокие позиции в поисковой выдаче. Путем множества манипуляций по работе с сайтом можно добиться хороших результатов в повышении конверсии сайтов. Современные системы по всевозможному анализу данных с помощью нейронных сетей могут сильно усовершенствовать работу над этой оптимизацией.

**Ключевые слова:** продвижение сайтов, поисковая оптимизация, нейронные сети, оптимизация кода, сверточные нейронные сети.

На сегодняшний день поисковые системы используют очень сложные алгоритмы поиска и упорядочивания огромных массивов информации для благоприятного пользовательского взаимодействия. Поисковая оптимизация (англ. Search engine optimization — SEO) является одним из ключевых факторов по объективной работе сайта и его поиску в поисковых системах.

Поисковая оптимизация, или SEO, включает в себя ряд комплексных мер по внешней и внутренней оптимизации сайта для того, чтобы повысить его в поисковых системах. Все меры, которые бизнес производит над сайтом, во благо повышения своего рейтинга в поисковых системах путем SEO оптимизации направлены на повышение трафика, проходящего через сайт, что, в свою очередь, повышает его выгодность с точки зрения бизнеса: чем больше пользователей будет на сайте, тем более выгодно можно продавать рекламу на этом сайте или продвигать какой-то контент, который приносит деньги [1].

Существует несколько способов оптимизации веб-сайтов. Есть внешняя оптимизация SEO, отвечающая за то, насколько наш сайт релевантен, основываясь на цитировании нашего сайта другими ресурсами. Внутренняя оптимизация является комплексом мер, направленным на

повышение именно внутреннего устройства сайта, его правильной структуры или же правильной семантической верстки [2].

При внутренней оптимизации сайта стоит следить за правильно построенной архитектурой сайта с точки зрения SEO, что требует кропотливого внимания [3]. Инструкции и правила описаны в документациях разных поисковиков, хоть они имеют и общие правила, но внутренний механизм работы самих поисковиков скрыт от всех, дабы не эксплуатировать его недостатки [4]. Из основных правил, с которым могут работать разработчики вебсайтов, является правильная верстка тегов сайта, так, например, в тег изображения рекомендуется включать альтернативный тег с описанием изображения, что, в свою очередь, влияет на повышения очков в ранжировании поисковых систем.

Помимо простого увеличения трафика за счет улучшения кодовой базы вебсайтов, следует отметить, что SEO являются практиками, которые помогают пользователям с ограниченными возможностями использовать тот или иной веб-ресурс. Под ограниченными возможностями понимаются не только проблемы пользователя и его особенного взаимодействия с сайтом, но также просто увеличение быстродействия может помочь людям, которые имеют ограничения по мощности используемых устройств, например, телефон, с которого будет происходить взаимодействие с сайтом. Кроме этого, улучшение сайта с точки зрения SEO, а именно - правильная настройка семантической верстки сайта, может помочь людям с ограниченными возможностями, которые используют какие-то дополнительные устройства или программы для пользовательского взаимодействия, ведь у них нет особого выбора. Так, например, можно выделить отдельный перечень программ для людей с ограничениями, такие программы называются “скринридеры” и позволяют использовать сайт в непривычном его понимании, а именно: дают возможность зачитывать вслух каждый элемент,

---

который встретится на странице [5]. Настройка сайта позволяет правильно переходить между его элементами и просто создает благоприятные условия использования ресурса. Так, если разработчик не озаботился тем, что элементы списка (метатег “li”) надо размещать внутри предназначенного для этого элемента (метатег “ul”), то браузер, “скринридеры” и просто поисковые системы могут неправильно обработать данный сайт, что, в свою очередь, может привести к ошибкам.

Стоит отметить, что можно делить существующие правила поисковых систем на разные категории. Один из самых популярных методов, когда мы делим правила на три класса (белый, серый и черный) в зависимости от цвета. Цвета, по сути, говорят, как мы будем повышать нашу SEO-оптимизацию. Белая оптимизация — это самая легальная работа над сайтом, она никак не связана с запрещенными техниками по продвижению сайта в поисковых системах, то есть эти методы никак не пытаются обмануть поисковые системы для продвижения ресурса. Остальные цвета, соответственно, говорят о том, насколько легальны наши методы по продвижению целевого вебсайта [6].

Современные методы машинного обучения могут сильно помочь в решении тривиальных задач, связанных с SEO-оптимизацией. Нейронные сети, которые помогают определять объекты на изображении, могут использоваться для того, чтобы заполнить недостающее описание тега изображений для поднятия очков SEO и, соответственно, помочь людям с ограниченными возможностями с меньшими трудностями использовать веб-ресурсы. А также генеративные нейронные сети для обработки машинописного языка могут помочь в более скрупулёзной обработке для SEO- оптимизации.

Для анализа изображений, на текущий момент, большего успеха добились сверточные нейронные сети, которые могут определять объекты,

---

которые находятся на картинке [7]. Такая нейронная сеть включает в себя несколько различных по своей сути и назначению слоев:

- Входной слой;
- Сверточный слой;
- Подвыборка;
- Обычные слои;
- Выходной слой.

Первые два слоя формируют вектор признаков многослойной сети, которая, в свою очередь, состоит из нескольких обычных слоев.

Все представленные слои служат для обучения нейронной сети, по сути, на выходе компьютер получает готовый алгоритм, по которому он может обрабатывать изображения и давать какой-то результат.

На входной слой программист подает изображения, которыми будет обучена нейронная сеть. Такие изображения должны иметь ряд ограничений, так, программист не может использовать изображения с огромным разрешением изображения из-за того, что вычислительные мощности не рассчитаны на обработку огромного массива данных, такие вычисления очень дорого обходятся не только с точки зрения сложности их вычисления, но также и их длительности [8]. Предварительно сжав изображение, можно добиться более быстрого обучения нашей нейронной сети.

На следующем этапе данные проходят через специальную функцию нормализации, которая позволяет выделить более четкие признаки. Все это происходит с помощью слоя свертки, который с помощью специальных ядер или распределяемых весов позволяет выделить более выраженные характеристики изображения [9].

Слой подвыборки отвечает за то, что исходный объем данных, приходящий с предыдущего слоя, сокращается. Это нужно потому, что следующий слой не требует огромного количества выявленных признаков, а

---

имеющиеся на данном слое данные уплотняются и выделяются более характерные признаки, такое сокращение данных устраняет проблему переобучаемости нейронной сети. Также данные проходят через функцию активации, которая, как раз - таки и отвечает за уменьшение количества выявляемых характеристик [10].

На одном из последних этапов выступает полносвязный слой, в свою очередь состоящий из множества обычных слоев. Назначение слоя — это моделирование сложной нелинейной функции. Оптимизируя эту функцию, можно улучшить качество распознавания изображения.

Выходной слой — это последний рубеж в нейронной сети между математикой и программистом, который связан со всеми обычными слоями из предыдущего слоя. В зависимости от решаемой задачи, выходной слой может представлять из себя от двух (бинарные задачи, например: определить, кошка на изображении или нет) до огромного количества выходных нейронов. Каждый нейрон отвечает за какой-то один объект из множества определяемых ответов.

Исходя из полученной нейронной сети, был разработан механизм определения объектов на изображении, для его последующего использования. Была использована библиотека Tensorflow для обучения и набор данных Flickr.

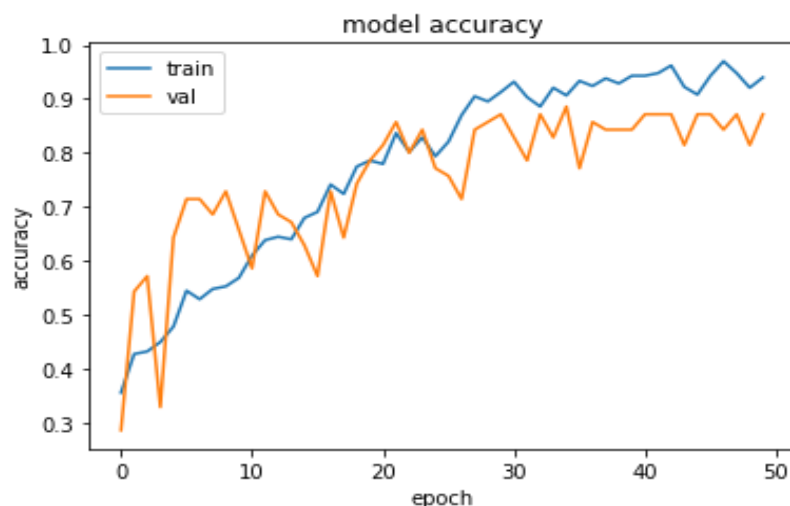


Рис 1. – Точность модели на тестовом наборе данных

Полученная модель и, впоследствии созданное к ней программное обеспечение, было использовано для создания сервиса по SEO.

Для реализации была использована библиотека Express.js с целью создания серверного сервиса по оптимизации html-верстки сайта, в частности, метатега `img`, для встраивания в него альтернативного тега. Примеры работы разработанного сервиса показаны ниже. Серверная часть не имела никакой информации об оригинальной фотографии, были отправлена только сама фотография и исходный html-код.

```
1 <body>
2     
3 </body>
```

Рис 2. – html-часть сайта до его оптимизации при помощи нейронной сети

```
1 <body>
2     
3 </body>
```

Рис 3. – html- часть сайта после его оптимизации сервиса.

Разработанный сервис, получив исходную фотографию, выдал к ней описание, основанное на построенной и обученной модели нейронной сети. Далее это полученное описание было сопоставлено с исходным html-кодом, который предварительно был декомпозирован на более элементарные единицы. И соответствующее описание было приписано к

соответствующему семантическому метатегу `img` и вписано в его параметр `alt`.

### **Выводы**

Путем использования современных технологий можно значительно улучшить семантику сайта для оптимизации SEO, использование нейронных сверточных нейронных сетей может помочь исключить существующие пробелы на вебсайтах, а также помочь людям с ограниченными возможностями в использовании сайтов.

### **Литература**

1. Бабаев А., Евдокимов Н., Боден М., Костин Е., Штарев А. // Раскрутка: секреты эффективного продвижения сайтов // Питер, 2013. С. 5-8.
  2. Иванов И.И // SEO: Поисковая оптимизация от А до Я // URL: [webbooks.com.ua/books-main/seo/ivanov-i-i-seo-poiskovaya-optimizaciya-ot/](http://webbooks.com.ua/books-main/seo/ivanov-i-i-seo-poiskovaya-optimizaciya-ot/)
  3. Киселев Ю.А. // Перспективы использования жанровой классификации Веб документов в поисковых системах // Инженерный вестник Дона, 2012, №4 URL: [ivdon.ru/ru/magazine/archive/n4p2y2012/1425](http://ivdon.ru/ru/magazine/archive/n4p2y2012/1425)
  4. Antoniou G., Harmelen F. // A Semantic Web Primer // The MIT Press Cambridge, Massachusetts London, England, 2012 URL: [prof.mau.ac.ir/images/Uploaded\\_files/A%20Semantic%20Web%20Primer-The%20MIT%20Press%20\(2012\)percentage5B7460174percentage5D.PDF](http://prof.mau.ac.ir/images/Uploaded_files/A%20Semantic%20Web%20Primer-The%20MIT%20Press%20(2012)percentage5B7460174percentage5D.PDF)
  5. Mukherjee S., Yang G., Tan W., Ramakrishnan I.V. // Automatic Discovery of Semantic Structures in HTML Documents // Department of Computer Science State University of New York at Stony Brook Stony Brook, NY 11794-4400. URL: [citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cdf917f46fcd89a85188c68347a60cf5aabce622](http://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cdf917f46fcd89a85188c68347a60cf5aabce622)
-



6. Enge E., Spencer S., Stricchiola J.C. // The Art of SEO mastering search engine optimization // O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. URL: [appcursosdegraca.s3.amazonaws.com/apostilas-en/information-technology/it-tools/the-art-of-seo-by-eric-and-jessie.pdf](https://appcursosdegraca.s3.amazonaws.com/apostilas-en/information-technology/it-tools/the-art-of-seo-by-eric-and-jessie.pdf)
  7. Rawat W., Wang Z. // Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review // Department of Electrical and Mining Engineering, University of South Africa, Florida 1710, South Africa URL: [researchgate.net/profile/Zenghui-Wang/publication/317496930\\_Deep\\_Convolutional\\_Neural\\_Networks\\_for\\_Image\\_Classification\\_A\\_Comprehensive\\_Review/links/59f814630f7e9b553ebefe27/Deep-Convolutional-Neural-Networks-for-Image-Classification-A-Comprehensive-Review.pdf](https://researchgate.net/profile/Zenghui-Wang/publication/317496930_Deep_Convolutional_Neural_Networks_for_Image_Classification_A_Comprehensive_Review/links/59f814630f7e9b553ebefe27/Deep-Convolutional-Neural-Networks-for-Image-Classification-A-Comprehensive-Review.pdf)
  8. Sharma N., Vibhor J., Anju M. // An Analysis Of Convolutional Neural Networks For Image Classification // Amity University Uttar Pradesh, Noida, India URL: [reader.elsevier.com/reader/sd/pii/S1877050918309335?token=E2E24B8C594642F00BA4943D703FFFE01F4B84C404806E2F99142EE6E9A4B0DDC806E987C4FCF6A12CA5EAE180BBB14B&originRegion=eu-west-1&originCreation=20230213204908](https://reader.elsevier.com/reader/sd/pii/S1877050918309335?token=E2E24B8C594642F00BA4943D703FFFE01F4B84C404806E2F99142EE6E9A4B0DDC806E987C4FCF6A12CA5EAE180BBB14B&originRegion=eu-west-1&originCreation=20230213204908)
  9. Arohan A., Koustav A., Abhishek S. // A Review of Convolutional Neural Networks // 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) URL: [researchgate.net/profile/ArohanAjit/publication/340968733\\_A\\_Review\\_of\\_Convolutional\\_Neural\\_Networks/links/5f9258cd92851c14bcded7f7/A-Review-of-Convolutional-Neural-Networks.pdf](https://researchgate.net/profile/ArohanAjit/publication/340968733_A_Review_of_Convolutional_Neural_Networks/links/5f9258cd92851c14bcded7f7/A-Review-of-Convolutional-Neural-Networks.pdf)
  10. Романов Д.Е. // Нейронные сети обратного распространения ошибки // Инженерный вестник Дона, 2009, №3. URL: [ivdon.ru/ru/magazine/archive/n3y2009/143](http://ivdon.ru/ru/magazine/archive/n3y2009/143)
-



## References

1. Babaev A., Evdokimov N., Bode M., Kostin E., Shtarev A. Raskrutka: cekrety effektivnogo prodvizheniya saytov [Promotion: the secrets of effective website promotion] Piter, 2013. Pp. 5-8.
  2. Ivanov I.I SEO: Poiskovaya optimizatsiya ot A do Ya [SEO: Search Engine Optimization from A to Z]. URL: [webbooks.com.ua/books-main/seo/ivanov-i-i-seo-poiskovaya-optimizaciya-ot/](http://webbooks.com.ua/books-main/seo/ivanov-i-i-seo-poiskovaya-optimizaciya-ot/)
  3. Kiselev Y. A. Inzhenernyj vestnik Dona, 2012, №4. URL: [ivdon.ru/ru/magazine/archive/n4p2y2012/1425](http://ivdon.ru/ru/magazine/archive/n4p2y2012/1425)
  4. Antoniou G., Harmelen F. A Semantic Web Primer The MIT Press Cambridge, Massachusetts London, England, 2012. URL: [prof.mau.ac.ir/images/Uploaded\\_files/A%20Semantic%20Web%20Primer-The%20MIT%20Press%20\(2012\)%5B7460174%5D.PDF](http://prof.mau.ac.ir/images/Uploaded_files/A%20Semantic%20Web%20Primer-The%20MIT%20Press%20(2012)%5B7460174%5D.PDF)
  5. Mukherjee S., Yang G., Tan W. Ramakrishnan I.V. Automatic Discovery of Semantic Structures in HTML Documents, Department of Computer Science State University of New York at Stony Brook Stony Brook, NY 11794-4400. URL: [citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cdf917f46fcd89a85188c68347a60cf5aabce622](http://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cdf917f46fcd89a85188c68347a60cf5aabce622)
  6. Enge E., Spencer S., Stricchiola J. C. The Art of SEO mastering search engine optimization, O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. URL: [appcursosdegraca.s3.amazonaws.com/apostilas-en/information-technology/it-tools/the-art-of-seo-by-eric-and-jessie.pdf](http://appcursosdegraca.s3.amazonaws.com/apostilas-en/information-technology/it-tools/the-art-of-seo-by-eric-and-jessie.pdf)
  7. Waseem R., Zenghui W. MIT Press, 2017, Issue 9. Department of Electrical and Mining Engineering, University of South Africa, Florida 1710, South Africa URL: [researchgate.net/profile/Zenghui-Wang-6/publication/317496930\\_Deep\\_Convolutional\\_Neural\\_Networks\\_for\\_Image\\_Classification\\_A\\_Comprehensive\\_Review/links/59f814630f7e9b553ebefe27/Deep-](http://researchgate.net/profile/Zenghui-Wang-6/publication/317496930_Deep_Convolutional_Neural_Networks_for_Image_Classification_A_Comprehensive_Review/links/59f814630f7e9b553ebefe27/Deep-)
-



Convolutional-Neural-Networks-for-Image-Classification-A-Comprehensive-Review.pdf

8. Neha S., Vibhor J., Anju M. Procedia Computer Science, Volume 132, 2018, Amity University Uttar Pradesh, Noida, India URL: [reader.elsevier.com/reader/sd/pii/S1877050918309335?token=E2E24B8C594642F00BA4943D703FFFE01F4B84C404806E2F99142EE6E9A4B0DDC806E987C4FCF6A12CA5EAE180BBB14B&originRegion=eu-west-1&originCreation=20230213204908](https://reader.elsevier.com/reader/sd/pii/S1877050918309335?token=E2E24B8C594642F00BA4943D703FFFE01F4B84C404806E2F99142EE6E9A4B0DDC806E987C4FCF6A12CA5EAE180BBB14B&originRegion=eu-west-1&originCreation=20230213204908)
9. Arohan A., Koustav A., Abhishek S. A Review of Convolutional Neural Networks, 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) URL: [researchgate.net/profile/Arohan-Ajit/publication/340968733\\_A\\_Review\\_of\\_Convolutional\\_Neural\\_Networks/links/5f9258cd92851c14bcded7f7/A-Review-of-Convolutional-Neural-Networks.pdf](https://researchgate.net/profile/Arohan-Ajit/publication/340968733_A_Review_of_Convolutional_Neural_Networks/links/5f9258cd92851c14bcded7f7/A-Review-of-Convolutional-Neural-Networks.pdf)
10. Romanov D.E. Inzhenernyj vestnik Dona, 2009, №3 URL: [ivdon.ru/ru/magazine/archive/n3y2009/143](http://ivdon.ru/ru/magazine/archive/n3y2009/143)