**Почему новостные заголовки**

Новостные заголовки в разных языках представляют собой явление, структурно и прагматически стоящее особняком от остальной языковой традиции. В них иначе кодируется грамматическое время, а прагматически они стремятся разделить с читателем его временной контекст. А в современном мире с распространяющимися за доли секунды через интернет новостями, заголовки играют все более сложную и важную роль.

Основная задача новостного заголовка – привлечь внимание читателя к событию и стать своего рода визуальным проводником к содержимому новостной статьи. Лучшего всего это заметно в социальных медиа, где в непрекращающемся потоке новой информации заголовки зачастую являются единственной видимой частью главного контекста. В отличие от чтения печатных изданий, при просмотре новостной ленты в интернете пользователь пробегается по заголовкам взглядом, выделяя из них ключевые слова и по ним определяя, что скорее всего будет сказано в статье и представляет ли она для него какой-либо интерес. В исследовании, проведенном издательством American Press Institute, исследователи выяснили, что как минимум 6 из 10 американцев не заглядывает в содержимое новостных статей, а только просматривает новостные заголовки, и исследователи предполагают, что на самом деле таких людей больше.

В автоматической обработке языка принято считать, что чем больше данных, тем лучше, чем длиннее тексты, тем больше шансов извлечь из них достаточное количество информации, чтобы научить машину понимать естественный язык. Мы же в свою очередь в данной работе хотим обратить внимание на то, что если человек может извлекать информацию из короткого заголовка и строить на его основании связи между темами, следить за их развитием во времени и отличать одну тему от другой, то заголовки содержат достаточное количество информации и стоит пытаться автоматически их обработать.

Переход на уровень заголовков может быть полезен при очень больших объемах обрабатываемой информации приходится отказываться от больших и эксплицитных сущностей в пользу небольших и емких. К примеру, ежедневно выходит около ххх новостей на английском языке. Разумеется, ежедневно в мире не происходит столько событий, многие новостные статьи являются повторами и пересказами той, что вышла первой, а также версиями изложения происходящих событий или содержания одной и той же пресс-конференции разными журналистами. На автоматическую обработку каждой такой статьи уйдет столько времени, что пока машина будет чистить корпус от повторов и извлекать информацию из каждой полной статьи, новость успеет устареть, а в некоторых областях торговли, например в высокочастотной торговле, существенна каждая тысячная доля секунды, потраченная на обработку новости.

Важность заголовков подтверждается также журналистами, существует немало руководств и учебников о том, как нужно писать новостные заголовки, что в них должно содержаться и в каком виде в них нужно преподносить информацию (Leigh, 2013). Некоторые исследования предлагают использовать новостные заголовки как средство анализа того, на что именно обращают внимание читатели, просматривая заголовки, на основе новостных критериев.

Автоматическое извлечение новостных критериев из заголовков может стать одним из центральных приспособлений во многих приложениях. Например, имея информацию о количестве просмотров статьи и выделив новостные критерии из заголовка, можно во‑первых, выяснить, какие именно из них заставили читателя открыть статью, во-вторых, сравнить, насколько аспекты, выделяемые в заголовке, отличаются от реальных вопросов, поднятых в статье, и насколько более или менее популярна за счет этого сама статья по сравнению с ее заголовком.

Как мы уже упоминали ранее, единицы меньшей длины требуют больших затрат на обработку содержания даже у человека. Для автоматической же обработки необходимы более совершенные алгоритмы извлечения темы, сущностей и новостных критериев. Автоматическая обработка осложняется еще и тем, что во многом успешное понимание заголовка сопряжено со знанием экстралингвистических фактов, которые далеко не всегда очевидны и не присутствуют в тезаурусах и словарях синонимов. Также не существует строгих определений новостных критериев, которые могут быть использованы при составлении текста заголовка, что обусловлено его природой. Строго говоря, не было проведено никаких исследований, которые показали бы, как связаны новостные критерии с той информацией, которая может быть автоматически извлечена из заголовка.