

NEWS Data Model

Team 6

February 14, 2018

NEWS data model solves the problem of randomly dispersed information over the Internet. It stores happenings of the real world.

- This valuable information has several parts such as a one line summary of the happening, called headline.
- To further elaborate the headline, a short description of it is also stored, which gives an insight into the article, without having to read it full.
- Detailed record of real world happening is called content.
- Along content our data model also saves the place where the event happened, it's date of publishing and the author of that article.

The data is saved in `.xml` format. It has the main root tag, named as `news`. It has sub-tags which are `current_affairs`, `sports` and `technology`, each of which stores a single article as `article` tag. To keep different attributes of an article separate, article tag further carries sub-tags which are `headline`, `description`, `content`, `place` and `author`. In those sub-tags of article, raw data is stored.

Hierarchy of NEWS data model:

- News
 - current_affairs
 - Article
 - Headline
 - Description
 - Author
 - Date
 - Place
 - Content

Sports

Article

- Headline
- Description
- Author
- Date
- Place
- Content

Technology

Article

- Headline
- Description
- Author
- Date
- Place
- Content

Here is the **schema** for the data model:

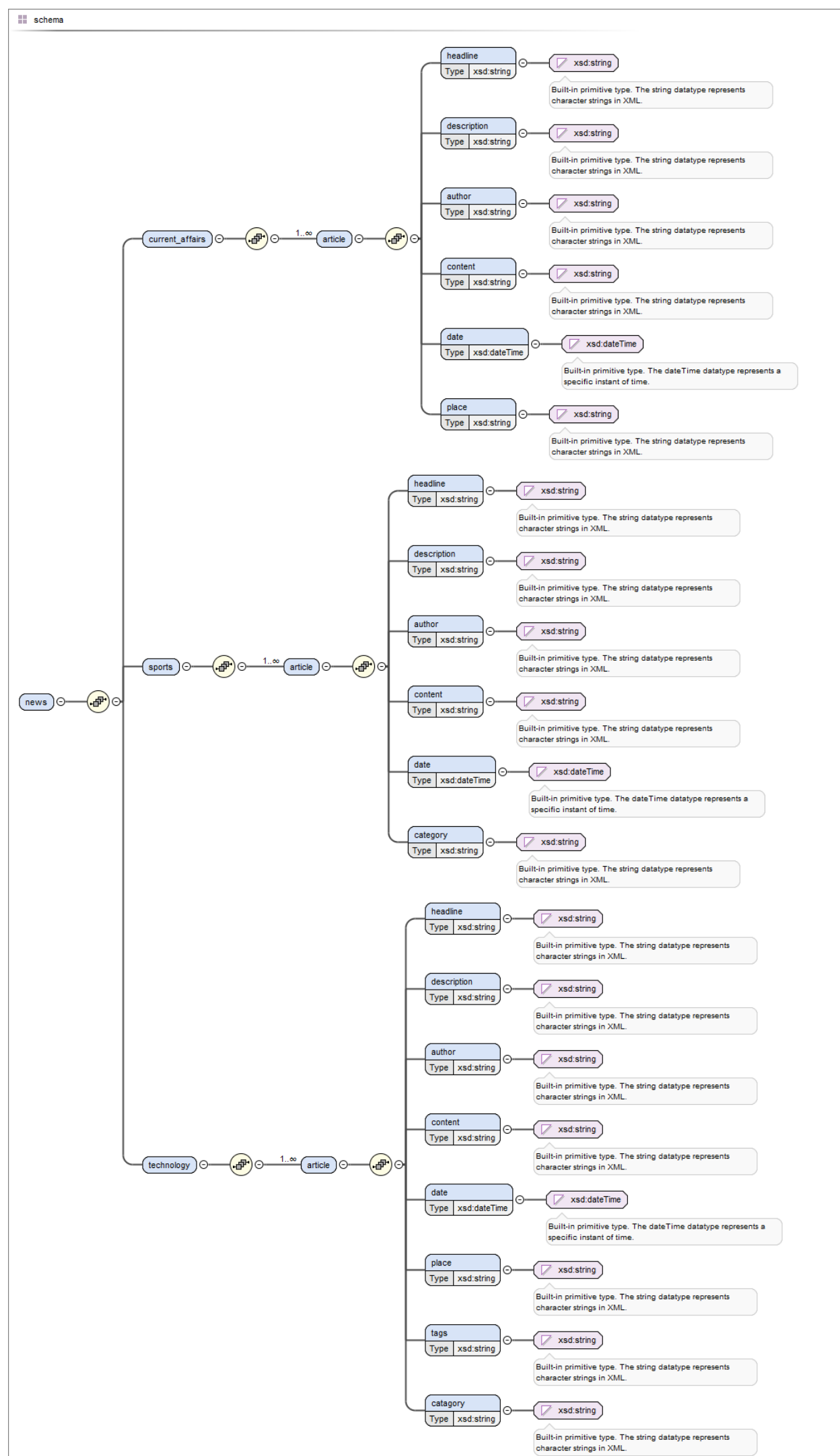


Figure 1: Schema

Tools used are:

- Stylus Studio
- **msxsl.exe** for command line execution
- BeautifulSoup4
- ET Tree Parser
- urllib library
- webbrowser library