

Statements about the Future – Automatically extracting Future-relevant Knowledge

APPLICATION

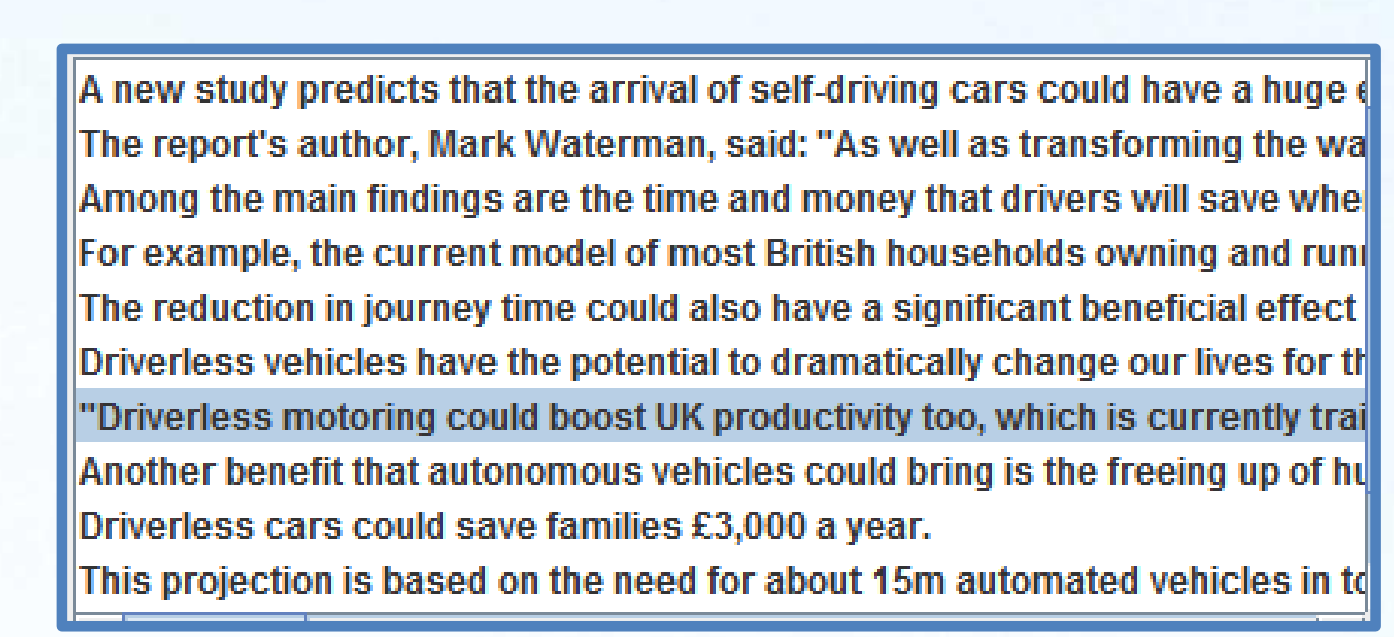
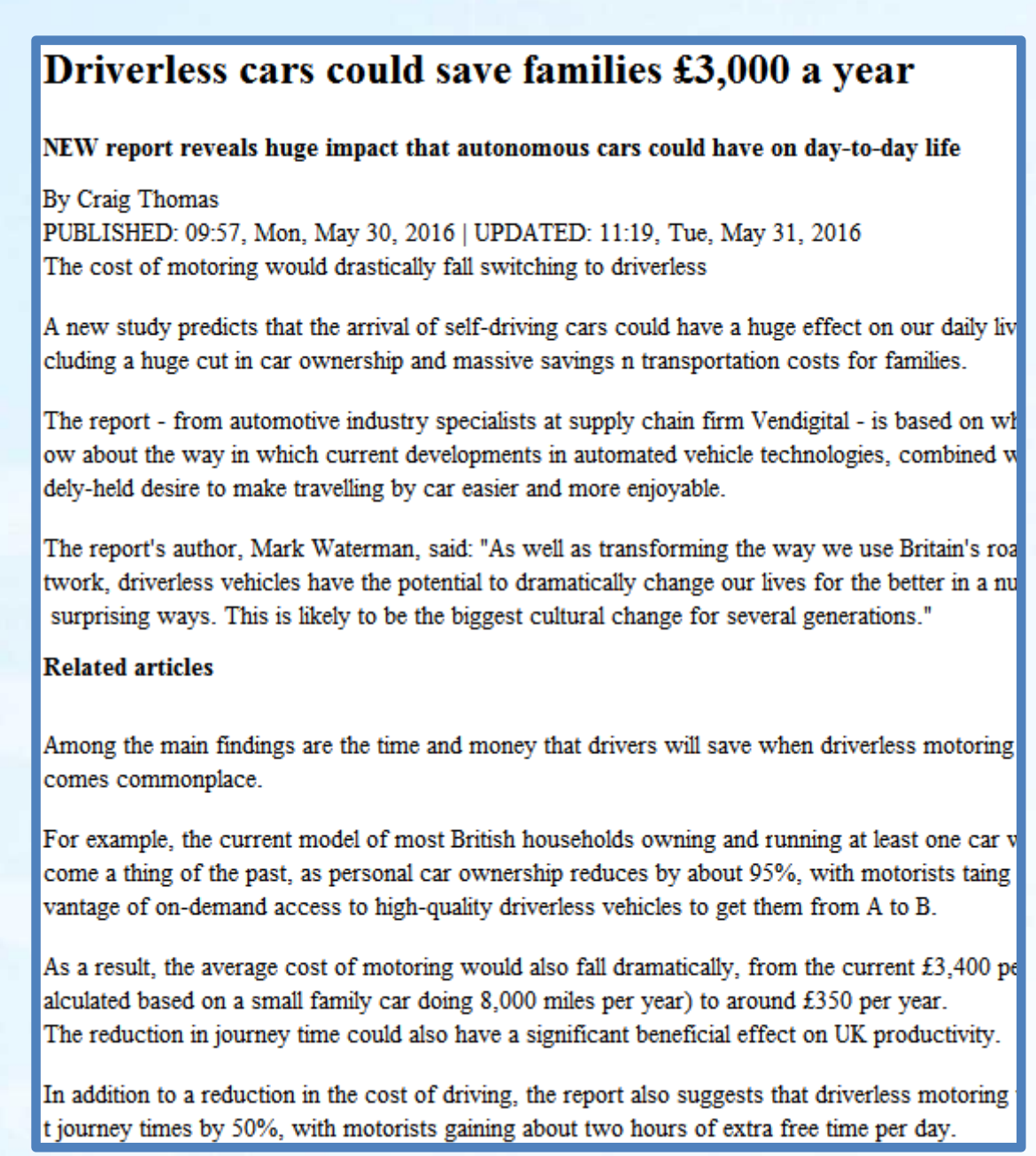
- Strategic foresight is the ability to create, maintain and use a high-quality, coherent and functional view of the future.
- One important activity inside Strategic Foresight is “**horizon scanning**”, the process of finding and interpreting information which has a strong influence on the future (e.g. trends, novelties).
- This activity can be made much more productive by combining text extraction algorithms and human quality control.

1 CLEAN TEXT EXTRACTION

- **Input: source URL**
- Source selection manual or automatic
- Download source documents
- Extract clean text from HTML, PDF, PPT (using structure, labels and text coherence)
- Extract document meta data: author, publication date, source, time range, keywords, organization, key people

2 FIRST ANALYSIS

- **Input: Clean Text**
- Integrated tokenizing and lexicon lookup
- Sentence splitting
- POS tagging
- Island Parsing
- Determine candidate sentences on checking «future-relatedness»



3 DEEP ANALYSIS

- **Input: Future-related Sentences**
- Deep (as necessary) syntactic analysis
- Condense sentence to shortest form → statement
- Remove context-dependent statements (co-references)
- Determine statement meta-data (start/end time, country/region, topics, type, organization, key person, quantities)
- Add relevancy metrics

4 PRODUCTION

- **Input: Future-related Statements**
- Manual post-processing (correction and deletion)
- Insert statements into database for search queries
- Prepare for web display
- Provide summaries and export formats

RESULTS

- In production since 2013
- Implemented in Java, processes ca. 20 web documents / second
- Part of general Foresight tool box
- Over 125 000 statements in database
- 1500 new statements per week
- General and client-specific services

