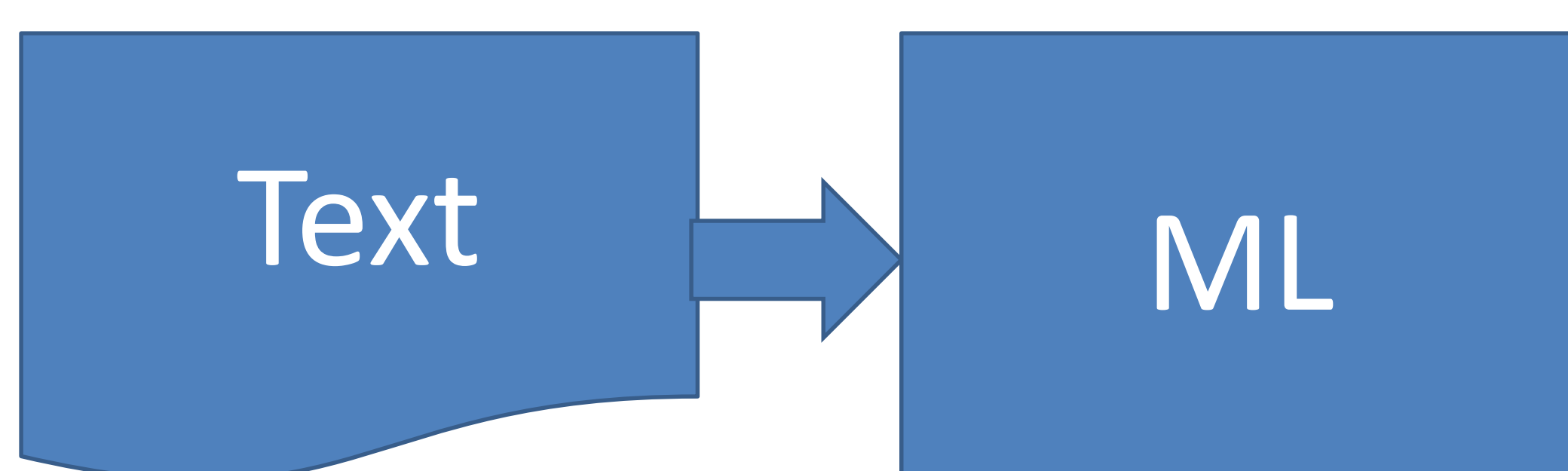


CERIAS Tech Report 2015-7
Meaning-Based Machine Learning
by Courtney Falk, Lauren Stuart
Center for Education and Research
Information Assurance and Security
Purdue University, West Lafayette, IN 47907-2086

Meaning-Based Machine Learning

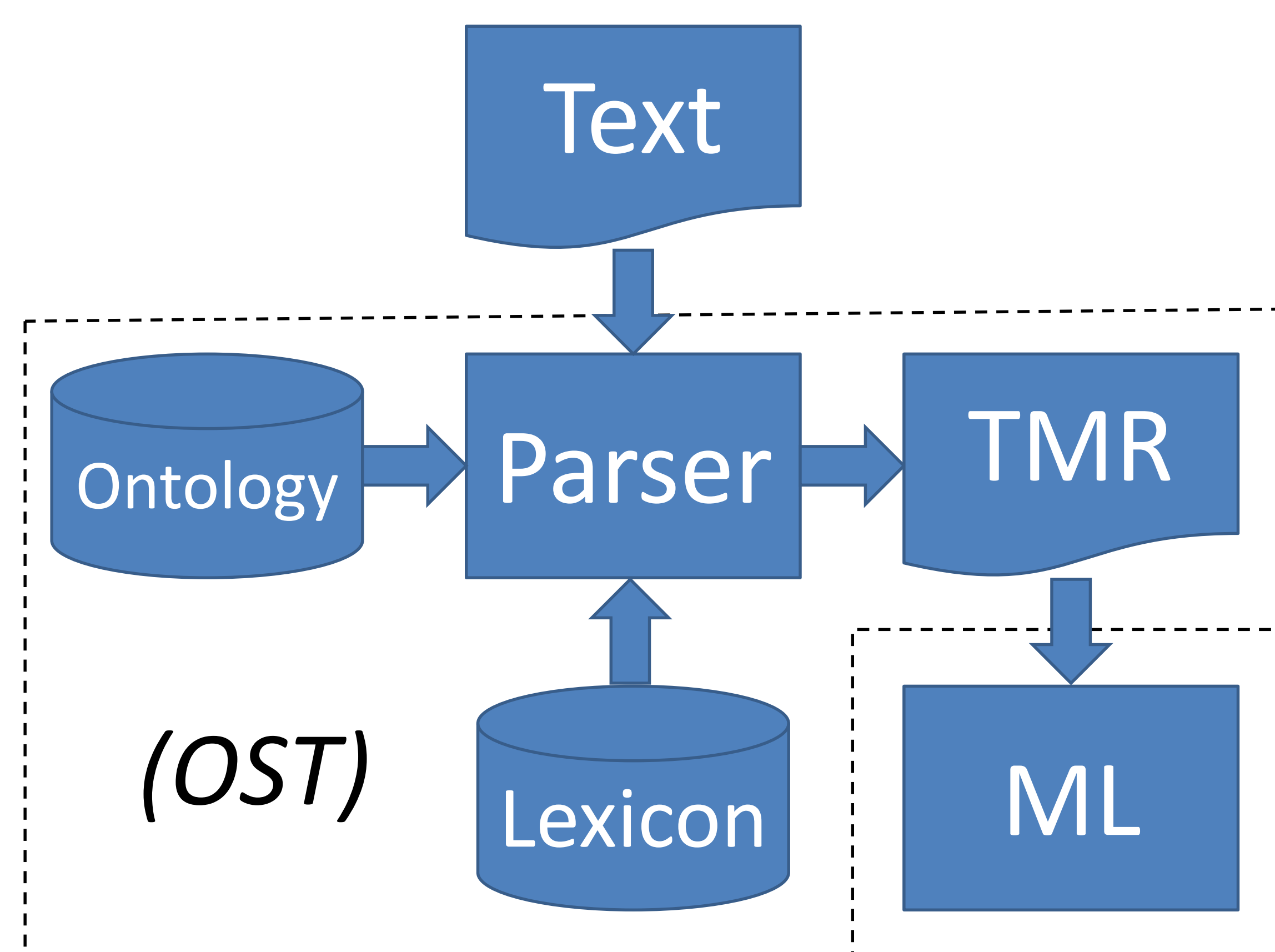
Courtney Falk and Lauren Stuart

Meaning-Based Machine Learning (MBML) is a research program intended to show how training machine learning (ML) algorithms on meaningful data produces more accurate results than that of using unstructured data.



Traditional statistical ML approaches try to find patterns or structures in unstructured data.

Ontological Semantics Technology (OST) is a knowledge-based natural language processing (NLP) system that supplies the meaning. Text meaning representations (TMRs) are the structured output of the OST system.



"President Hollande visited the American capital today."

(VISIT
(AGENT
(HUMAN-117
(PLAYS-ROLE (PRESIDENT-42))))
(DESTINATION
(CITY-13
(CAPITAL-OF (NATION-1))))))

ML algorithms are trained on the TMR structures generated by the OST parser. These algorithms identify patterns in the meaning of the texts that isn't immediately apparent.

<VISIT, AGENT, HUMAN-117>
<HUMAN-117, PLAYS-ROLE, PRESIDENT-42>
<VISIT, DESTINATION, CITY-13>
<CITY-13, CAPITAL-OF, NATION-1>

Possible applications include:

Phishing detection Stylometry Anonymization
Search engines Summarization Watermarking