



Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics and
Electrical Engineering
Department of Computer and Information Science

Mikel Arizaleta

Structured data extraction: separating content from noise on news websites

Trondheim, June 2009