

# ASSISTED STRUCTURED AUTHORING USING CONDITIONAL RANDOM FIELDS

By Bert Willems (@devatwork)

# Structured authoring examples

Wolfgang Amadeus Mozart (27 January 1756 – 5 December 1791), baptised as Johannes Chrysostomus Wolfgangus Theophilus Mozart, was a prolific and influential composer of the classical era.

Aspirin comes as 300mg tablets. The usual dose is 1 or 2 tablets every 4 to 6 hours.

# XML to the rescue?

**Yes, but SMEs are not happy to do additional work!**

# Why not ~~Spell~~Markup Check?

It is:

- Helpful
- Not intrusive

# Machine learning: Intro

- Ability to learn without being explicitly programmed
- Natural Language Processing
- Text Classification

# Machine learning: Anatomy

- Input
- Model & Weights
- Loss function
- Backpropagation

# Tying it all together

- Conditional Random Fields
- Text classification
- Label to XML mapping

Nobody is perfect



## Recognized Bibliographic Citation

*Kristel Pappel*, *Von der Wandertruppe zum ständigen Theater*, in:  
*Musikgeschichte in Mittel- und Osteuropa*, Chemnitz 1998, 3-13

Please check and correct the classifications:

*Von der Wandertruppe zum ständigen Theater*

as article title - 98% ▾

*Musikgeschichte in Mittel- und Osteuropa*

as journal title - 95% ^

journal title - 95% ✓

series title - 50%

Apply markup

# State of the implementation

- Conditional Random Field: done
- UX Design: meh
- Integration: !time
- Next step: user testing

**FONTOXML**

Thank you!