

Semesterarbeit “Adaptive Weighting of Filters in Spamato ”

Dieses Dokument gibt den Rahmen der Semesterarbeit von Dennis Rietmann im SS 2006 vor. Abweichungen oder Änderungen sind in gegenseitiger Absprache möglich. Diese Arbeit ist für 150 Arbeitsstunden ausgelegt.

Arbeitsplan:

Gegenstand dieser Arbeit ist die automatische Gewichtung von Spamfilter-Resultaten im Spamato Systems. Im Rahmen dieser Arbeit sollen die folgenden Aufgaben behandelt werden:

- 1) **Einarbeitung in das Spamato System.** Geklärt werden sollen Fragen der Art: Wie ist Spamato aufgebaut? Wie funktioniert das Plugin-System und wie lässt es sich erweitern? Wie lassen sich die Spamfilter-Resultate erfassen? Wie können „Decisions“ erstellt werden?
- 2) **Erarbeitung und Analyse verschiedener Gewichtungs-Algorithmen:**
 - Eine genauere Betrachtung der derzeitigen Spamfilter-Resultate soll Aufschluss darüber geben, wie diese am besten (niedrige False/Positive- und False/Negative-Raten) zu gewichten sind. Wie können Gewichtungs-Algorithmen einfach miteinander verglichen werden? Kann dafür eine Art „Kennzahl“ berechnet werden?
 - Es sollen verschiedene Ideen zur Gewichtung der Resultate betrachtet und bewertet werden. Ansatzpunkte sind z.B. Neuronale Netze oder Genetische Algorithmen. Aber auch eine „einfache“ Gewichtung anhand der False/Positive- und False/Negative-Raten kann herangezogen werden.
 - Wie verändert sich das Verhalten, wenn sich Resultate „plötzlich“ verändern (etwa wenn der Bayesianato nach dem Lernen aktiv wird)? Was passiert, wenn neue, bisher unbekannte Filter hinzukommen oder vorhandene gelöscht werden? Wie kann man sich dynamisch auf solche Veränderungen einstellen?
 - Wie können einfache „Regeln“ (Ruleminator) und „Subfilter-Resultate“ (Razor) miteinbezogen werden? Können verschiedene Einstellungen berücksichtigt werden?
- 3) **Implementierung:**
 - Implementierung des „Evaluations-Frameworks“.
 - Implementierung einer oder mehrerer Gewichtungs-Algorithmen als Plugin für Spamato.
 - Test und Evaluation im praktischen Alltag.
- 4) **Erstellung eines kurzen Berichts** (10 Seiten) über die Arbeit und die Resultate. Der Bericht soll abschliessend auch eine kritische Beurteilung der eigenen Arbeit enthalten.

Allgemeines:

- Selbstständiges Arbeiten ist Voraussetzung.
- Nach der Arbeit ist eine Abschluss-Präsentation vorgesehen.
- Treffen können mit Keno jederzeit „bei Bedarf“ stattfinden.

Kontaktpersonen:

- | | |
|----------------------|----------------------------|
| 1. Keno Albrecht | kenoa@tik.ee.ethz.ch |
| 2. Roger Wattenhofer | wattenhofer@tik.ee.ethz.ch |