

LFE Medieninformatik - Projektarbeit

Flock Mail

Michael Weiler

Agenda

- **Einleitung**
 - Spam Statistiken & Filter
 - Schwarmintelligenz
 - Verwandte Arbeiten
- **Design**
 - Motivation
 - GUI
 - Visualisierung von eMails
- **Implementierung**
 - Ähnlichkeitsberechnung zwischen eMails
 - Bewegungsprinzip nach Craig Reynolds
 - UML Diagramm und Quellcodeausschnitt
- **Abschluß**
 - Live Demo
 - Ausblick

Einleitung - Spam

- 2006 - 55 Milliarden pro Tag (ironport.com)
- 2007 - 100 Milliarden pro Tag (spamunit.com)
- 97% aller Mails sind Spam (BBC News Website)

1420 in 30 Tagen (gmail) - ca. 50 pro Tag

Tanon Tyra	Events listed below - if this e-mail does not display correctly, view it in your browser July 2009 contact us visit ...
Watches Omegas	Handbags - Hello, old chap! Take a look - http://hereskye.com/ Good bye
Jesse	Koennen Sie es ihr hart geben? - Hallo! Sie haben Probleme beim Sex? Die Hektik macht Sie fertig und hat auch noch negative Folgen ...
Casino Yal	Das beste Angebot überhaupt. - Sie wurden ausgewaehlt einen atemberaubenden Willkommensbonus im Privilege Club Casino in Empfang ...
Neues Jahr - 30% Billiger	Es ist Unglaublich Sicher Hier zu Bestellen - Mit der Nummer D33W-3351 zahlen Sie 30% weniger fuer alle unsere Produkte. Nur Legale Software ...
tuolloin1973	For weiler - If you cannot see the pictures and links below, please click here to view them. If nothing seen ...
Milford Ruqboes	Maybe cinema? - To view a web version of this message, click here. BUY NOW WITH 80% OFF More Items On SALE This ...
Schifo	» Rock-hard erections Without the Blrue Pzill - Rock-hrad erections Without the Bltue Pill www.med95.net . Workers fiind unopened letter too ...
Dr. Michelle Thoma	Abnehmen, schnell und erfolgreich - Richtig abnehmen kann heut zu tage unschwer sein und sogar schon Vergnuegen bringen. Probieren ...
weiler	Your iTunes Account #248582 - Tell a friend Download latest version See this email as a webpage Hello! Shipped Privately And ...

Einleitung - Spam Filter

Listenbasierend



- Blacklist
- Whitelist

Regelbasierend



- Score = Spamwahrscheinlichkeit
- HTML Formatierungen
 - gefälschte SMTP Komponenten
 - bestimmte Schlagworte

Bayes-Filter

- Selbstlernendes System
- Anfangstraining erforderlich

Kollaboratives Filtern

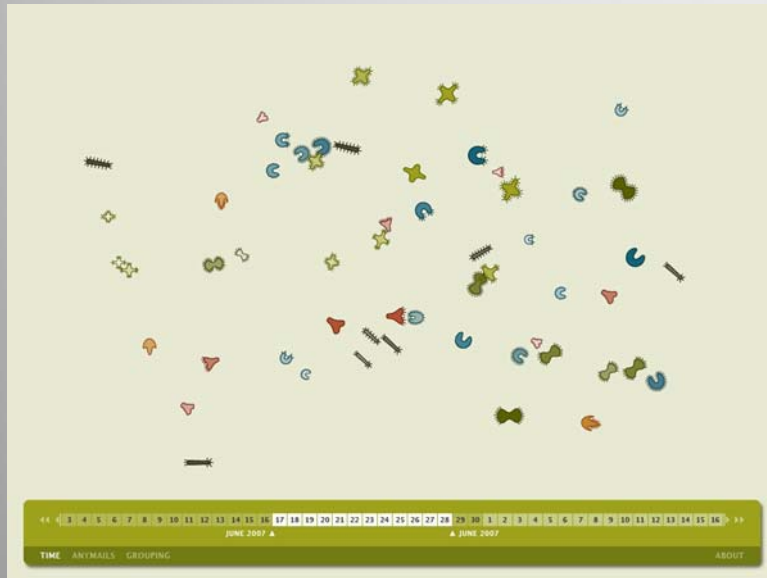
- Serverseitige Beobachtung
- Vergleich neuer Mails mit Spam-Mails
- Automatische Mechanismen und Benutzer gesteuert

Einleitung - Schwarmintelligenz

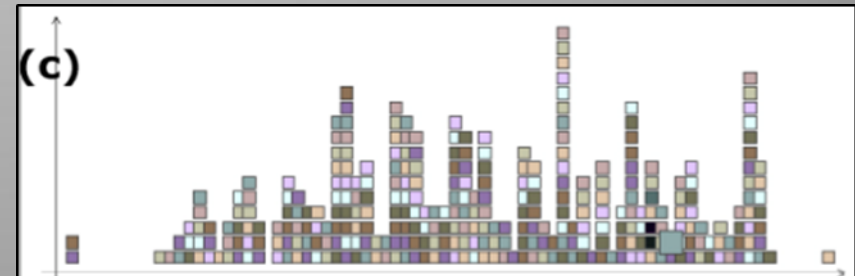
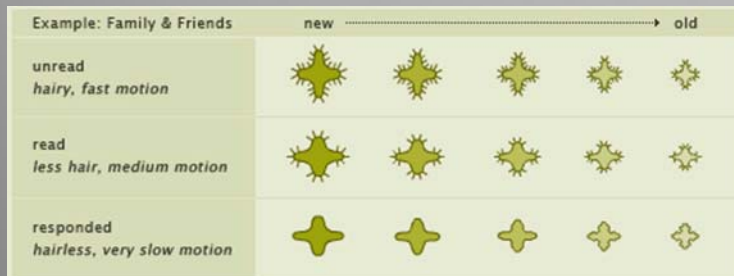
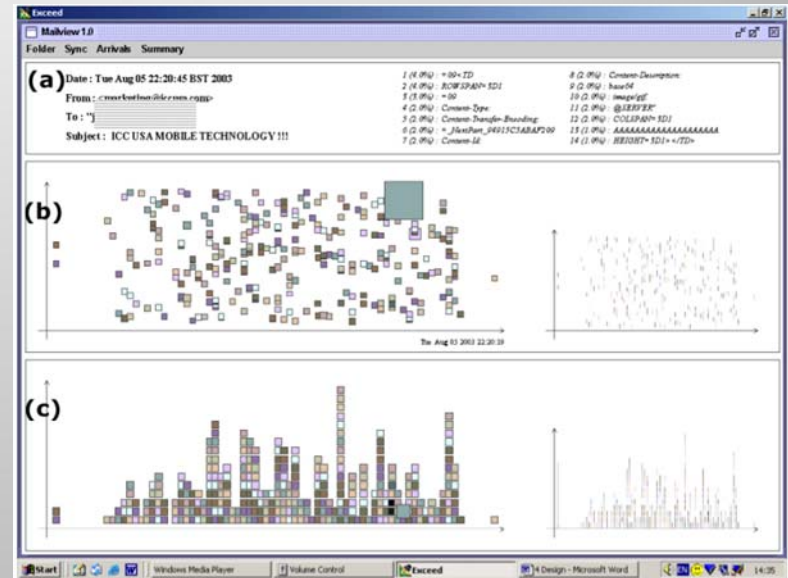
- Forschungsfeld der Künstlichen Intelligenz
- Basierend auf Agententechnologie
Autonom, Proaktiv & Reaktiv, Sozial Lernfähig
- Problemlösung durch natürliches Vorbild:
Insekten, Vogel- oder Fischschwärme

Einleitung – Verwandte Arbeiten

Anymails (Carolin Horn, Florian Jenett)

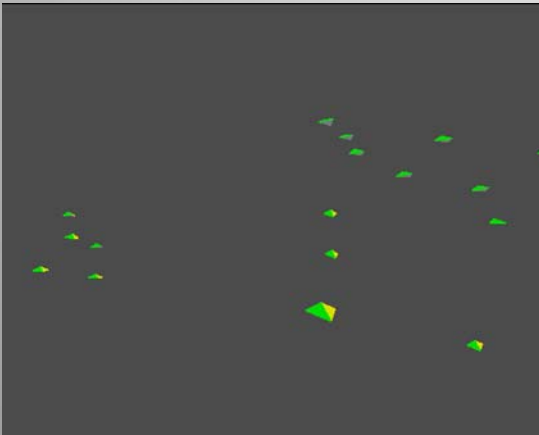


MailVilew (S. Frau, J.C. Roberts, and N. Boukhelifa)

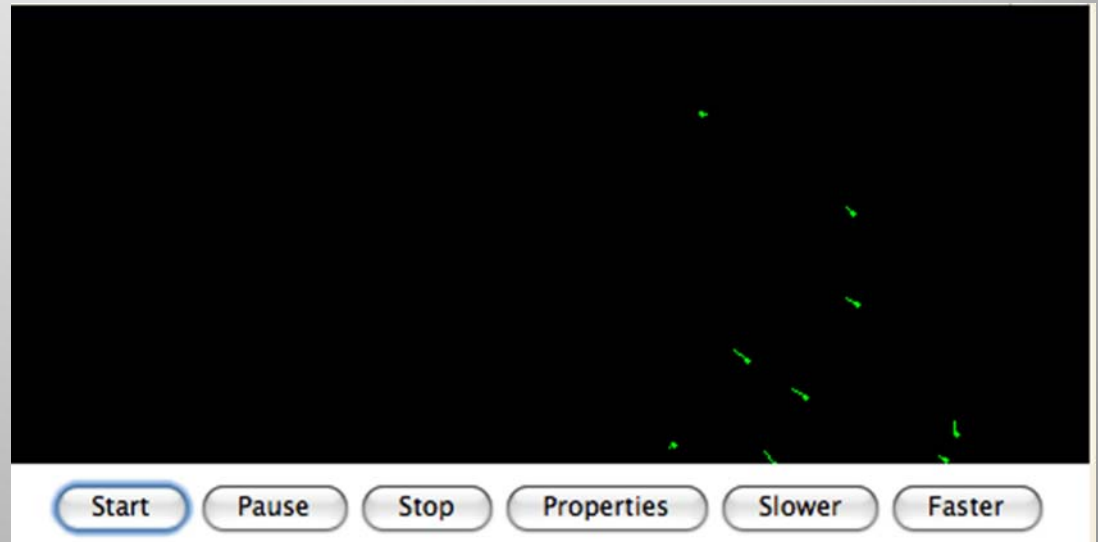


Einleitung – Boids Applets

Craig Reynolds 1986



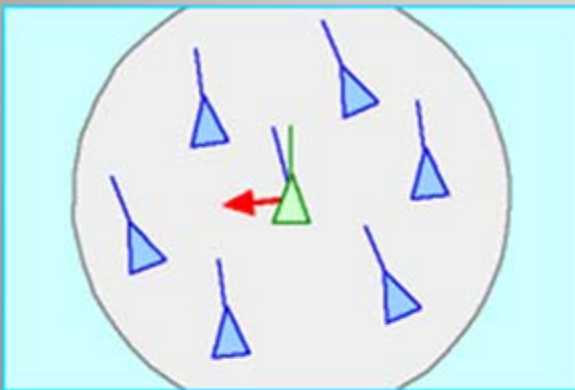
Floys, An Experiment In Java Alife (A. Dolan)



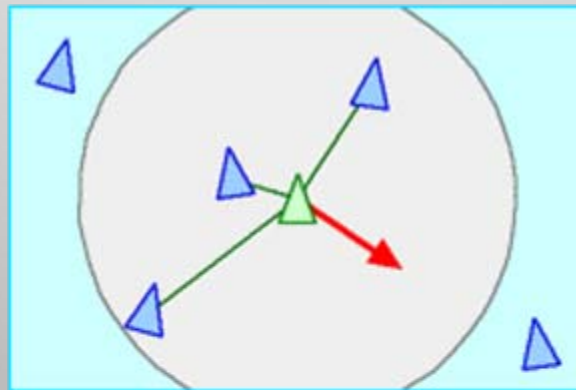
Einleitung - Boids

(Craig Reynolds 1986)

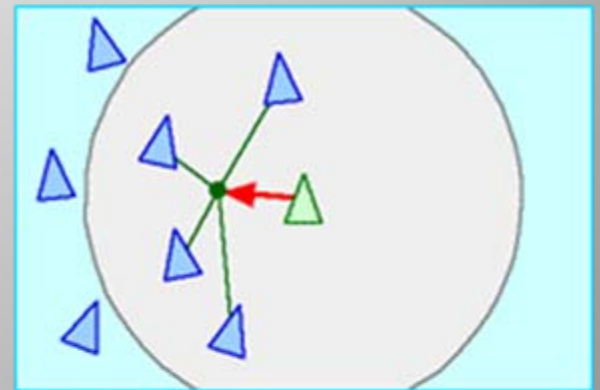
Angleichung



Separation



Zusammenhalt



Einleitung – Vorteile des Boid Prinzips

Zuweisung: Boid \Leftrightarrow eMail

- eMails i.A. massenhaft (gut durch Boids darstellbar)
- **Lokalitätsvorteil** (Sichtradius)
- Effiziente natürliche **“Sortierung”** bzw. **Anordnung**
- Einfaches **Anpassen des Verhaltens** (durch Hinzufügen von neuen Bewegungsregeln)

Einleitung - Motivation

Vorteile gegenüber herkömmlichen Methoden (Outlook, gmail):

- Spamfreie Mailbox bleibt spamfrei durch **selbstorganisation**
- Entdecken von **Tendenzen**
- Steigerung der Inbox-**Übersichtlichkeit**
- Automatische **Clusterbildung** (z.B. Gruppierung nach Autor)
- **Interessantes Verhalten** neuer eMails
- **Spaßfaktor** steigern beim täglichen Abrufen der eMails

Design – Was ist Flockmail?

- Stand-Alone Email-Client (Java 1.6 & Eclipse SDK 3.4.2)
- Verbindung von eMails und Boids
- Versuch Spam durch Boid Verhalten zu erkennen (selbstständiges Clustering)
- “Lebendige” Mailbox
- Andere Sicht auf eMails
- Interaktion des Benutzers (Auswahl bestimmter eMail Cluster, eMails schreiben)

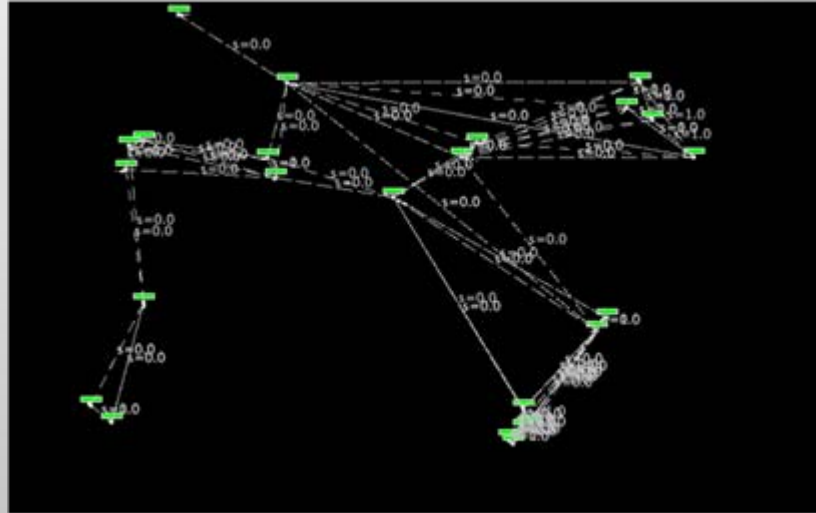
Design – Settings Dialog

The image shows a 'Properties' dialog box with the following fields and controls:

host	<input type="text" value="imap.gmail.com"/>
user	<input type="text" value="helpmeapp"/>
password	<input type="password" value="••••••••"/>
port	<input type="text" value="995"/>
mails from date	<input type="text" value="2009/01/01"/>
maximum mails	<input type="text" value="25"/>
account type	<input checked="" type="radio"/> IMAP <input type="radio"/> POP

Buttons:

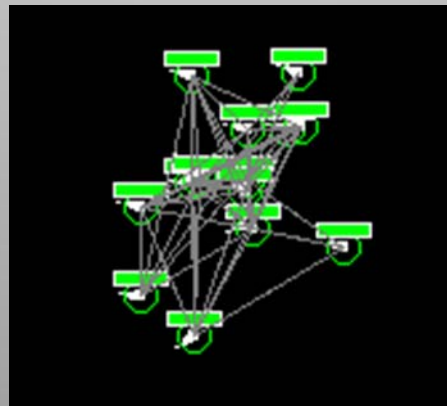
Design – Visualisierung im Detail



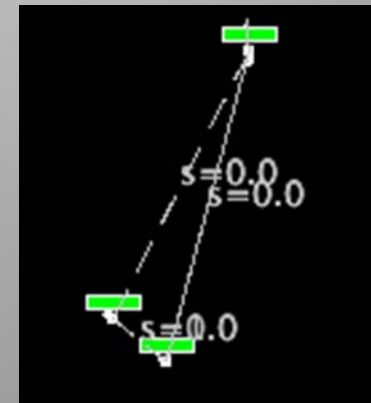
Macro Boids



Boid Cluster



Ähnlichkeitsmessung



Implementierung- Ähnlichkeitsfunktionen

AuthorSimilarity (1 falls Autor gleich, 0 sonst)

CosineSimilarity

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

EuclideanDistance

$$\text{euc}(q, r) = \left(\sum_y (q(y) - r(y))^2 \right)^{1/2}$$

MongeElkan

$$\text{match}(A, B) = \frac{1}{|A|} \sum_{i=1}^{|A|} \max_{j=1}^{|B|} \text{match}(A_i, B_j).$$



Aufwendige Berechnung => Speicherung in Ähnlichkeits Matrix

Implementierung - Berechnungsfunktionen

$$\bar{x}_{arithm} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\bar{x}_{arithm} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

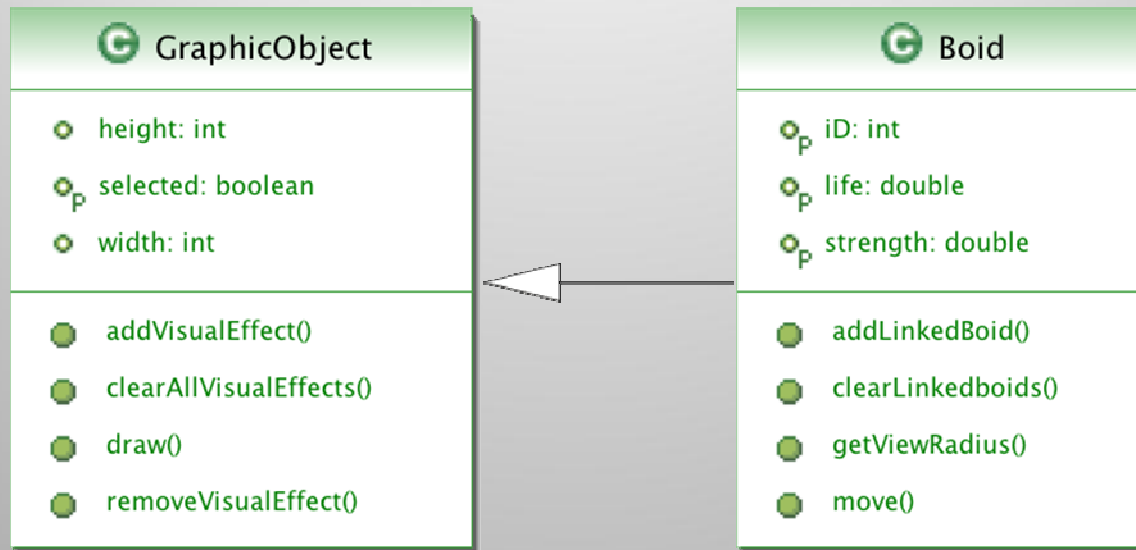
$$\max(x_1, x_2, \dots, x_n)$$

$$\max(x_1, x_2, \dots, x_n)$$

$$\bar{x}_{geom} = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

Implementierung – Boid Klasse

vereinfachte UML Diagramme



=> Darstellung eines Boid auf der Zeichenfläche (Paint Thread)

Implementierung – Boid.move()

```
public void move() {
    PointDouble sum_v = new PointDouble();
    for (BoidMoveRule rule : rules) {
        PointDouble delta =
            rule.getVelocityVector(this);
        sum_v.shift(delta);
    }
    velocity.x = velocity.x * 0.4 + sum_v.x * 0.4;
    velocity.y = velocity.y * 0.4 + sum_v.y * 0.4;
    position.shift(velocity);
}
```

- Bewegungsregeln: Richtungsänderung von 0 bis 1 für je x und y
- Aufsummierung in Vektor sum_v
- Bewegungsvektor wird entsprechend sum_v angepasst (40%)

Implementierung – Modularisierung durch Listen

Boid Klasse

```
private ArrayList<BoidMoveRule> rules;  
private ArrayList<SimilarityMetric> metrics;
```

- Reynolds Bewegungsregeln (Separation, Angleichung, Zusammenhalt)
- Im Sichtbereich bleiben
- Meiden ungleicher Boids
- Makroboidgenerierung

VisualEffect Klasse

```
private static ArrayList<VisualEffect> effects;
```

- Ähnlichkeitsmessung
- Makroboid Visualisierung
- Lebensbalkenanzeige

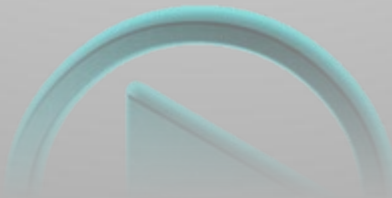
Implementierung - Einschränkungen

- Performance: Visualisierung vieler eMails und komplexe Ähnlichkeitsregeln in Verbindung mit textreichen eMails problematisch
- HTML: Ähnlichkeitsvergleich lediglich auf blankem Text
- Dateianhänge: werden nicht berücksichtigt
- Kampfmodus, Lebensenergie: derzeit nicht implementiert
- Benutzerinteraktion beschränkt auf: Clusterauswahl, eMailversand und -empfang.

Ausblick

- Verknüpfung von bereits erfolgreich angewandten **Spamfilter-Methoden** mit FlockMail
- Erweiterung um **Neuronale Netze** und **Evolutionäre Algorithmen**
- Implementierung eines Kampf-Modus (Spam vs. Non-Spam)
- Erweiterte Benutzerinteraktion (Anpassen von Regelparametern)
- Erweiterung auf 3D
- Optionales Plugin (z.B. Thunderbird)

Live Demo



Danke für Ihre Aufmerksamkeit!

