



Erfassung von Internetfragebögen mit SAS

Wolf F. Lesener
Humboldt-Universität zu Berlin
Rechenzentrum

Potenziale von Datenerhebungen in Netzen¹



- Zeit- und Kostenersparnis
- Mehrfachverwendung
- Teilnehmerrekrutierung
- Einbindung in weitere Abläufe

¹) D. Janetzko „Statistische Anwendungen im Internet“ Addison-Wesley 1999

Probleme bei Datenerhebungen in Netzen¹



- **Fehlende Repräsentativität**

- *Als repräsentativ wird eine Stichprobe angesehen, wenn jede Beobachtungseinheit der Grundgesamtheit die gleiche Chance hatte, in die Stichprobe einzugehen (Zufallsstichprobe) oder wenn in Abhängigkeit von der Zusammensetzung der Grundgesamtheit die Stichprobe entsprechend anteilig zusammengesetzt wird.¹ S. 141*

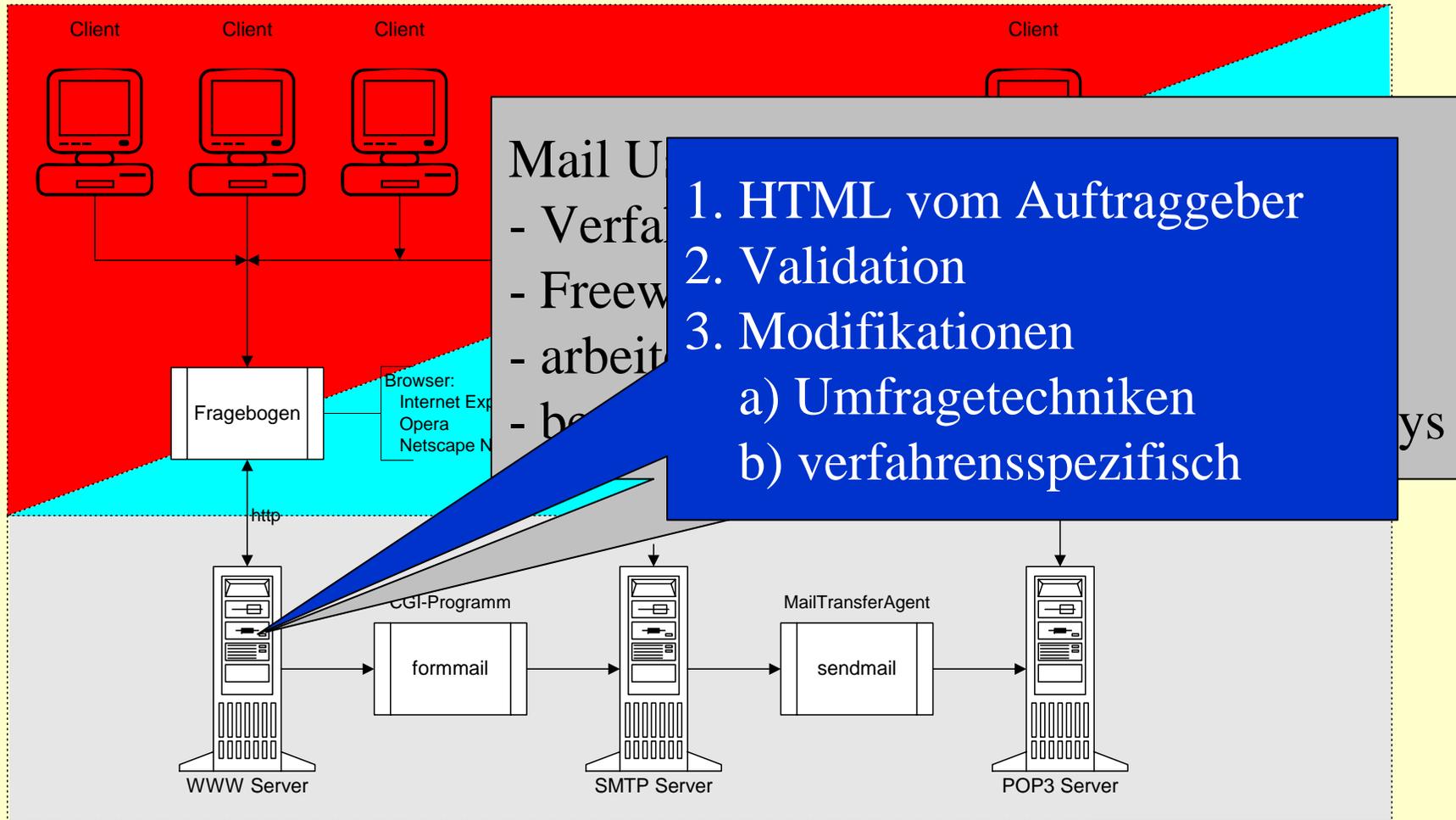
¹) D. Janetzko „Statistische Anwendungen im Internet“ Addison-Wesley 1999

Voraussetzung



Es ist gründlich zu prüfen, ob eine Umfrage über das Internet das geeignete Arbeitsmittel ist.

Informationsfluß



¹ im Zusammenhang mit Datenerfassung von Internetfragebögen nebensächlich

Erforderliche Kenntnisse



Computer-Sprachen	Auftrag- geber	Entwickler (RZ)	Proband
HTML	gut	sehr gut	
JavaScript		gut	
Cascaded Style Sheets	gut	gut	
Perl		gering	
Base SAS Language	gering	sehr gut	

sehr gut = selbständig kreativ; gut = an Beispielen orientiert; gering = black-box-Anwendung

Der Kern: HTML Formular



- FORM - Formular
- INPUT - Eingabefeld gemäß TYPE=
- SELECT - Auswahlliste
- OPTION - Optionen zu SELECT
- TEXTAREA - mehrzeiliger Text
- Script Programmierung (JavaScript, VB, ...) zur Auswertung von Ereignisattributen

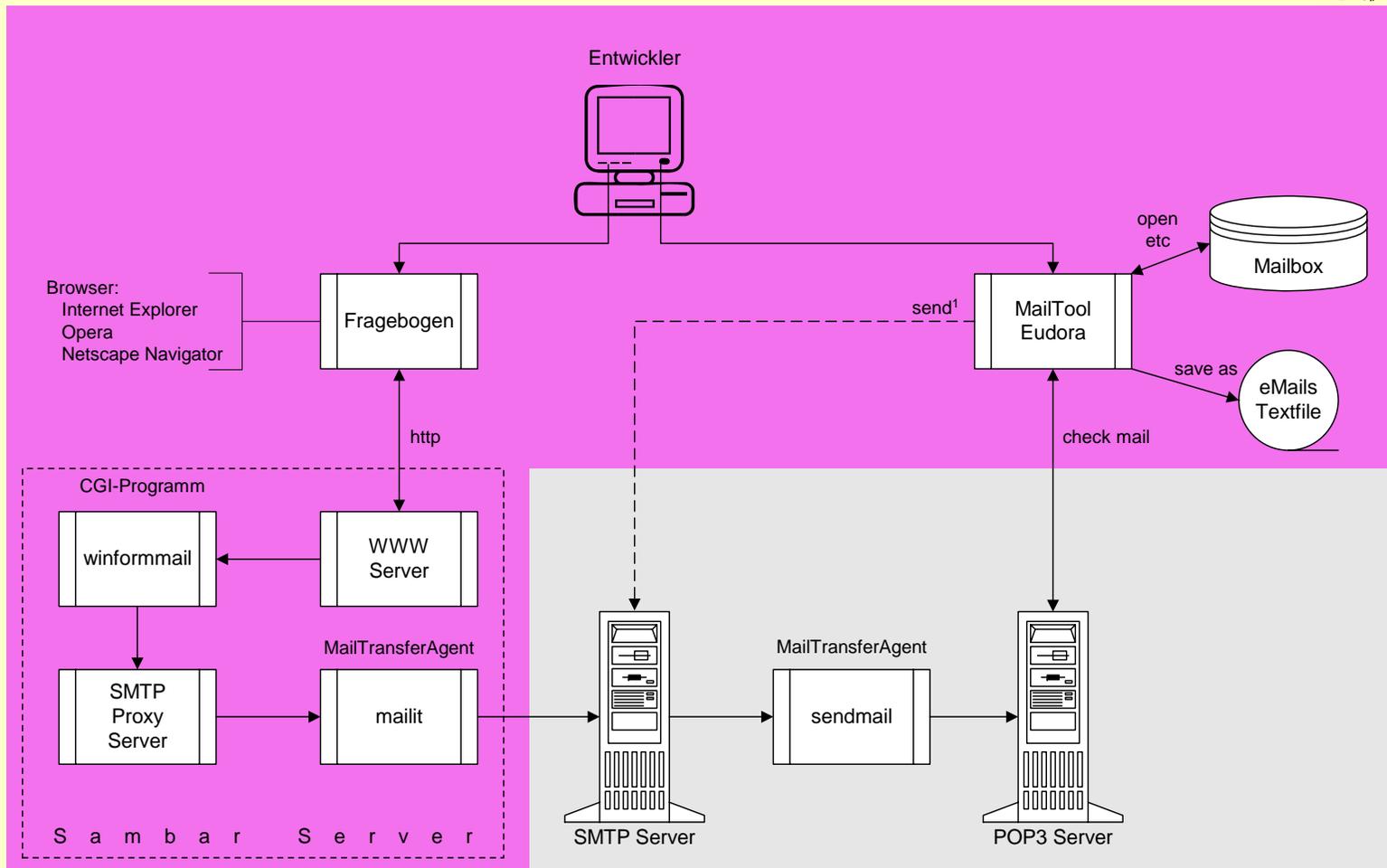
Umfragetechniken²



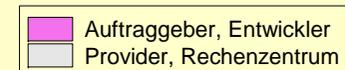
- Zufallsauswahl der Stichprobe (nth viz)
- Kontrolle des Non-Response
- Plausibilitäts- und Konsistenzkontrollen
- One-Screen-One-Question
- Filterführung
- Fortschrittsanzeige

²) <http://www.demotopia.de> „Methoden der Online-Befragung“

Die Testumgebung



¹ im Zusammenhang mit Datenerfassung von Internetfragebögen nebensächlich



Software



- Sambar Server 5.0
ActivePerl Build 522
- FormMail 1.6 (1.9) (-> winformmail.pl)
- SAS 8.2
- Eudora 5.1
- Internet Explorer 6.0
Netscape Navigator 6.2
Opera 6.0

Datenaggregation



Jede Datenzeile enthält einige technisch bedingte zusätzliche Variablen.

Eine aktuelle Datentabelle wird über die Felddauer mit PROC APPEND fortgeschrieben.

Evtl. mehrfach erfaßte Fragebögen werden nach Ablauf der Feld-

Die Anzahl der zurückgesandten Fragebögen begrenzt den Stichprobenumfang n

genügt, sie einmal permanent zu speichern.



Generieren eines eMail-Interpreters



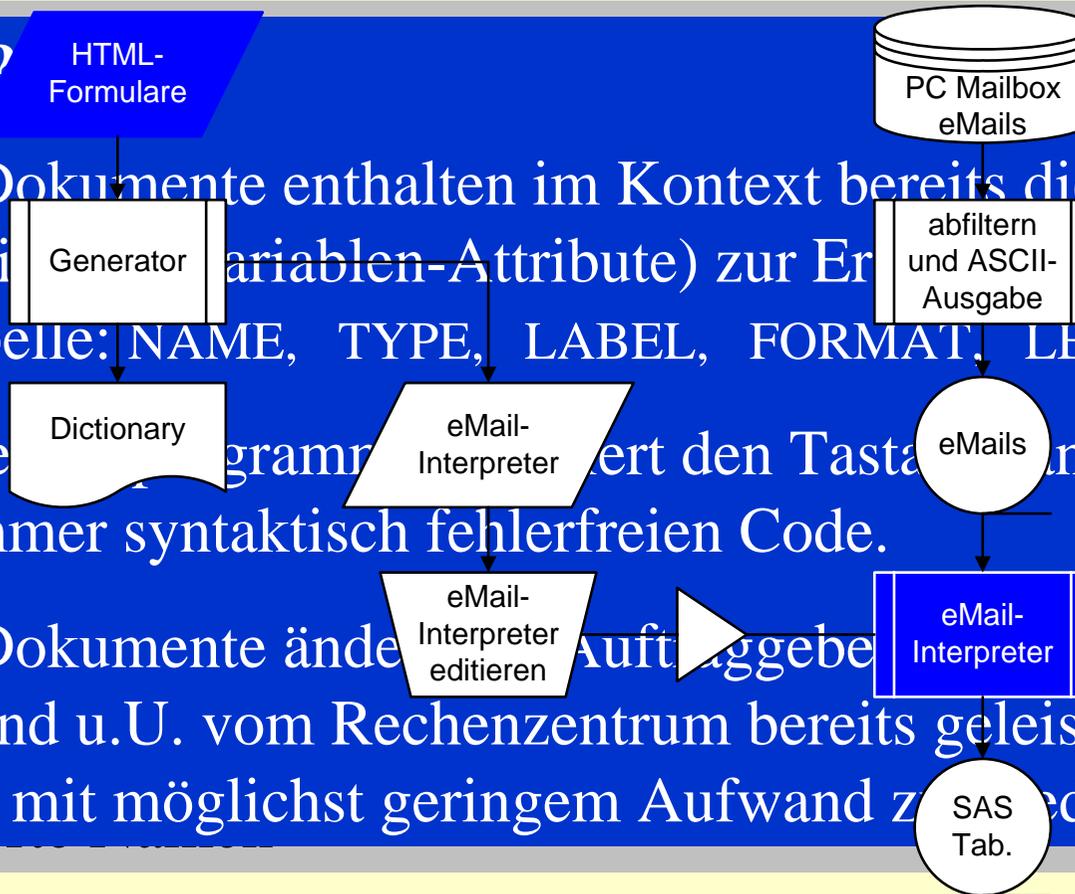
Warum?

HTML-
Formulare

HTML Dokumente enthalten im Kontext bereits die Meta-
Informationen (Variablen-Attribute) zur Erzeugung der SAS
Datentabelle: NAME, TYPE, LABEL, FORMAT, LENGTH

Ein Generator programmiert den Tastaturcode und
liefert immer syntaktisch fehlerfreien Code.

HTML Dokumente ändern sich häufig,
d.h. es sind u.U. vom Rechenzentrum bereits geleistete
Arbeiten mit möglichst geringem Aufwand zu wiederholen



Ein Beispiel



- Sambar Server starten
 - Fragebogen (Internet Explorer)
 - eingehende eMails (Eudora)
 - Generator starten
 - Das Interpreter-Programm
 - Editierten Interpreter starten
 - Die SAS Datentabelle
- } SAS

