



Martin Spickermann

Einführung in die Techniken wissenschaftlicher Umfragen

Empirische Forschungsmethoden in der Softwaretechnik
Blockseminar im Sommersemester 2004

FU Berlin • FB Informatik
AG Software Engineering

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 1/29



Einführung 1

Umfragen sind eine weit verbreitete Methode zur Erhebung von:

- Meinungen
- Trends
- Zusammenhängen
- Erfahrungen

Mögliche Vorteile gegenüber anderen Untersuchungsmethoden (wie z.B. Experimenten):

- Liefern einen guten Überblick
- Sind meist kostengünstiger
- Auf verschiedenste Gebiete anwendbar

Aber Achtung: Durchführung von Umfragen ist nicht trivial.

- Ergebnisse sind nur brauchbar, wenn bestimmte Regeln bei der Erstellung, Sammlung und Analyse der Daten eingehalten werden.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 2/29



Einführung 2

Bestandteile einer Umfrage sind nicht nur die Fragen als solche, sondern auch:

- Zielfestlegung (Hypothese, Fragestellung)
- Auswahlkriterien für die Teilnehmer
- Kritische Auswahl der Fragen
- Pilot-Test oder Focus-Test
- Test auf Gültigkeit und Verlässlichkeit
- Durchführung
- Dokumentierung
- Datenauswertung
- Veröffentlichung der Ergebnisse

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 3/29



Grundlage dieses Vortrags

6-teilige Artikelserie in der „Software Engineering Notes“ (2001 – 2003).

- SEN ist eine Publikation der Sigsoft (Special Interest Group on Software Engineering)
- Sigsoft ist eine Gruppe der acm (Association for Computing Machinery), der ältesten IT-Vereinigung (seit 1947)



Barbara A. Kitchenham PhD
Professor of Quantitative Software Engineering
Keele University, Staffordshire, UK
Mitglied der Royal Statistical Society

Dr. Shari Lawrence Pfleeger PhD
President of Systems/Software Inc .
Arlington, Virginia, USA
IEEE Mitglied

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 4/29



Überblick

Überblick über diesen Vortrag

- Forschungsziele
- Design und Arten von Umfragen
- Zielgruppen
- Fragenkatalog
- Zuverlässigkeit (reliability)
- Gültigkeit (validity)
- Stichprobe (sampling)
- Rücklaufquote
- Auswertung

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 5/29



Beschreibung des Forschungsziels

Was wollen wir erfahren?

- Formulierung einer Hypothese oder zu prüfenden Aussage.
- Die Aussage oder Frage muss klar und messbar sein, denn an ihr werden folgende Punkte bestimmt:
 - Wer wird interviewt?
 - Wie wird interviewt?
 - Welche Informationen müssen gesammelt werden?
 - Wie sehen die Fragen aus?
 - Abhängige und unabhängige Variablen?
 - Ist eine Umfrage überhaupt die richtige Forschungsmethode?

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 6/29



Design einer Umfrage 1

Beschreibende Umfragen (descriptive)

- Momentaufnahme (cross sectional)
 - Schwerpunkt auf dem Zeitpunkt
 - Umfrage an mehrere Personen zu einem ganz bestimmten Zeitpunkt.
z.B. „Woran arbeiten Sie gerade jetzt?“
- Gruppenumfrage (cohort)
 - Schwerpunkt auf der befragten Gruppe
 - Umfrage zu einem Thema an einen bestimmten Personenkreis
z.B. an alle Programmierer einer Firma.
- Fallkontrolliert (case control)
 - Schwerpunkt auf dem Thema
 - Meist ein Rückblick auf gemachte Erfahrungen mit einem bestimmten Thema: „Was halten Sie von Mustern im Softwaredesign?“

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 7/29



Design einer Umfrage 2

Experimentelle Umfrage (experimental)

- Gleichzeitige Kontrollstudien (concurrent control studies)
 - in denen die Teilnehmer zufällig gruppiert werden:
 - In denen die Teilnehmer nicht zufällig gruppiert werden:
- Vergleichende Kontrollstudien
 - Mit dem Teilnehmer selbst (self control studies)
 - Mit früheren Untersuchungen (historical control studies)
- Kombinationen
 - z.B. 1.2 + 2.1 : Teilnehmende und Nichtteilnehmende einer Fortbildung werden vor und nach der Veranstaltung befragt.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 8/29

 Arten von Umfragen

- Einzelbetreut (supervised)
 - Telefonische oder persönliche Interviews unter vier Augen
- Gruppenbetreut (administered)
 - Gruppenbefragungen, in denen auch Verständnisfragen erklärt werden.
- Beaufsichtigt (semi-administered)
 - Gruppenbefragungen, in denen die Fragen (häufig Fragebögen) vorgestellt und erläutert werden. Bei der eigentlichen Beantwortung der Fragen wird aber keine Hilfestellung mehr gegeben.
- Unbeaufsichtigt (unsupervised)
 - Umfragen im Internet oder in Zeitungen.

Komplexe Fragen
Teuer
Beeinflussung möglich
Hohe Rücklaufquote

Einfache Fragen
Günstig
Kaum Beeinflussung
Niedrige Rücklaufquote

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 9/29

 Zielgruppe

Festlegung und Bestimmung einer genau definierten Zielgruppe, denn nur so kann man folgende Punkte beurteilen oder entwerfen:

- Kann die Gruppe die Fragen überhaupt beantworten und zur Verifizierung der Hypothese beitragen?
- Zu benutzende Sprache in den Fragen.
- Größe der Gruppe, damit zu wählende Stichprobengröße (sample) und berechenbare Rückläuferquote.
- Art der Zusammensetzung der Gruppe, damit die Stichprobe repräsentativ ist.
- Kommunikationskanal um die Teilnehmer zu erreichen

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 10/29

 Vorbereitung zum Entwurf des Fragenkatalogs

Vor Beginn der Erstellung eines Fragenkatalogs sollte man alte Studien zu ähnlichen Themen lesen,

- damit man sie nicht versehentlich wiederholt.
- um aus deren Fehlern oder Lücken zu lernen.
- denn bekannte und bewährte Methoden sind aus Gründen der Vergleichbarkeit (Gültigkeit und Verlässlichkeit) vorzuziehen.

Bei Übernahme von Teilbereichen unbedingt:

- Neu auf Gültigkeit und Verlässlichkeit prüfen.
- Neuer Pilot-Test.

Doch nicht immer kann man deren Methoden oder Fragen übernehmen:

- Unterschiedlich geartete Gruppe oder andere Umfrageart.
- Umfang nicht passend.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 11/29

 Fragenentwurf 1

Tipps zum Entwurf der Fragen:

- Die Fragen sollten immer einen erkennbaren Zusammenhang zur Hypothese haben.
- Frage muss sprachlich als auch sachlich überhaupt beantwortet werden können.
- Keine Entweder/Oder-Fragen, keine negierten Fragen, sondern klare vollständige Fragen ohne jeglichen Spielraum zur Interpretation.
- Nicht allgemein fragen, sondern Zeitrahmen festlegen.
- Nicht zu viele Fragen stellen

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 12/29



Fragenentwurf 2

Tipps zum Entwurf der Fragen:

- Standardisierte Antworten
- Immer auch ein Feld „Weiß nicht“ oder „neutral“ anbieten.
- Bei eigenen Skalen empfehlen sich ca. 5-7 Stufen.
- Sorgfältige Auswahl der Fragenart:
 - geschlossene Fragen
 - offene Fragen
- Neutrale nicht wertende Fragen, klare unvoreingenommene Einleitung.
- Vorsicht mit zu persönlichen Fragen (z.B. Einkommen)

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 13/29



Vortest / Vorstudie

Nachdem der Fragenkatalog fertig ist, sollte er geprüft werden.

z.B. durch:

- Eine Gruppe von Fachleuten zu dem Thema, die nicht an der Erstellung der Fragen mitgewirkt haben.
- Eine Vorstudie (pilot studies) mit ausgewählten Teilnehmern (focus group), die eine repräsentative Auswahl aus der späteren Stichprobe darstellen.

Dadurch findet man:

- Fehler oder Unstimmigkeiten
- Unnötige, überflüssige Fragen
- Erste Hinweise zur Brauchbarkeit und Reproduzierbarkeit
- Feedback auch in Hinblick auf unbeabsichtigte Beeinflussung oder Fehler im Design.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 14/29



Zuverlässigkeit (reliability)

Um aussagekräftige Ergebnisse zu bekommen muss sicher gestellt sein, dass die Antworten zuverlässig (reproduzierbar) sind. Mögliche Arten von Zuverlässigkeiten, die geprüft werden können:

- Wiederholungstest (test-retest (intra observer) reliability)
 - Die gleiche Person wiederholt zu einem späteren Zeitpunkt den Test, bei einer Korrelation mindestens 0,7 ist von einer Zuverlässigkeit auszugehen.
- Veränderte Fassung (alternate form reliability)
 - Wiederholungstest, mit umgeformten und umgruppierten Fragen.
- Innere Konsistenz (internal consistency)
 - Auftauchen von umgeformten Fragen in der gleichen Umfrage mit dem Ziel die Gleichartigkeit der Antworten zu belegen.
- Reproduzierbarkeit (inter-observer (inter-rater) reliability)
 - Vergleich der Antworten verschiedener Probanden auf die gleichen Fragen. (siehe El Emam / SPICE)

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 15/29



Gültigkeit (validity) 1

Weiterhin muss geprüft werden, ob die Umfrage aussagen kann, was erkundet werden soll. Möglichkeiten des Tests:

- **Scheinbare Gültigkeit (face validity):** Durchsicht der Fragen von einem Amateur. Nicht sehr vielversprechend, da zu subjektiv.
- **Inhaltliche Gültigkeit (content validity):** Durchsicht der Fragen von Experten des Themas. Auch keine statistisch belegte Garantie.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 16/29

 Gültigkeit (validity) 2

Weitere Gültigkeitskriterien

- Vergleichsgültigkeit (criterion validity):** Vergleich der neuen Umfrage mit älteren, ähnlich gearteten Umfragen.
- Konstruktive Gültigkeit (construct validity):** Quasi ein Beweis über die gestellten Fragen.

Pfleeger behauptet, dass in der Mehrzahl aller Umfragen zum Thema Software Engineering die Gültigkeit nicht hinreichend belegt seien.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 17/29

 Stichprobe (sample)

Wenn es sich nicht gerade um äußerst kleine Zielgruppen handelt, ist es nicht sinnvoll alle Personen tatsächlich zu befragen.

- zu teuer.
- administrativer Aufwand zu hoch.
- unmöglich

Man wählt also eine geeignete, repräsentative Teilmenge aus.

- Repräsentativ bedeutet, dass alle relevanten Eigenschaften der Zielgruppe (z.B. Alter, Geschlecht, Parteizugehörigkeit) im gleichen Verhältnis auch in der Stichprobe auftreten.
- Nur wenn die Stichprobe repräsentativ ist, lässt sich das Ergebnis der Stichprobe auf die gesamte Gruppe übertragen.

Die Größe der Zielgruppe muss bekannt sein,

- damit anhand der Rücklaufquote festgestellt werden kann, ob die resultierenden Antworten überhaupt repräsentativ sein können.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 18/29

 Größe der Stichprobe

- Zu große Stichproben sind zu teuer in der Durchführung und fördern die Genauigkeit nur unwesentlich.
- Zu kleine Stichproben lassen keine Verallgemeinerung auf die gesamte Gruppe zu.

Berechnung der optimalen Größe des Samples mit folgender Formel:

$$\text{sampleSize} = \left[\frac{(z_a - z_b) * \sigma}{\mu_1 - \mu_2} \right]^2$$

wobei

$\mu_1 - \mu_2$ = erwartete Größendifferenz (aus pilot - test)
 σ = Standardabweichung (aus pilot - test)
 z_a = ist das obere Ende in der Standardnormalverteilung zu α . α ist Wahrscheinlichkeit eines Type I Fehlers.
 z_b = ist das obere Ende in der Standardnormalverteilung zu β . β ist Wahrscheinlichkeit eines Type II Fehlers.
Bsp. $z_a = 1,96$ wenn $\alpha = 0,05$.
Bsp. $z_b = -0,84$ wenn $\beta = 0,20$.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 19/29

 Auswahl der Stichprobe 1

Ein Sample ist statistisch betrachtet nur gültig, wenn es zufällig entstanden ist. Folgende zufallsbasiierende Methoden (probabilistic sampling methods) sind erlaubt:

- Einfacher Zufall (simple random sample)**
- Geschichteter Zufall (stratified random sample)**
- Systematischer Zufall (systematic sampling)**
- Zufällige Bündelauswahl (cluster-based sampling)**

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 20/29



Auswahl der Stichprobe 2

Dennoch kann es notwendig sein, nicht-zufallsbasierte Methoden zu benutzen.
Achtung: Die Ergebnisse sind dann ggf. nicht mehr repräsentativ.

- **Bequemlichkeitsauswahl (convenience sampling):** Man nimmt jeden, der freiwillig mitmachen möchte (z.B. Website oder Manager). Gefahr von Verfälschung und „freeping“ sehr groß.
- **Schneeballauswahl (snowball sampling):** Man wählt einen Teilnehmer bewusst aus, um lässt ihn dann andere vorschlagen. (Bsp. Hacker, man kennt halt nicht so viele).
- **Quoten Auswahl (quota sampling):** Bildung von Untergruppen, dann weiter mit 1 oder 2. Gegenstück zum geschichteten Zufall.
- **Focus groups:** persönliche Auswahl von Personen, die repräsentativ erscheinen.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 21/29



Verzerrungen bei der Durchführung

Verschiedene Faktoren können das Ergebnis beeinflussen.

- Verhalten des Interviewers während eines Interviews oder auch die Veränderung des Verhaltens im Verlauf der gesamten Studie.
- Gruppenverhalten der Befragten, auch wenn sie nicht direkt kontrolliert werden.
- Sympathie zwischen Interviewten und Interviewer
- Umfragen zu bestimmten Zeiten
- Ehrlichkeit der Befragten
- Umfragen unter Druck oder in unangenehmer Atmosphäre
- Fehlende Antwortmöglichkeiten oder Überschneidung der Antworten

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 22/29



Rücklaufquoten

Um sicher zu stellen, dass die Antworten, die man auf seine Fragen erhält, wirklich noch repräsentativ sind, ist es wichtig, die Quote der abgegebenen Antworten im Verhältnis zu den angesprochenen Personen zu kennen.

- Schwierig, wenn man die möglichen Teilnehmer gar nicht kennt.
- Nachforschen, warum Personen nicht geantwortet haben.
- Quote und den Umgang damit in der Auswertung darstellen.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 23/29



Tipps für höhere Rücklaufquoten

- Klare formulierte Fragen.
- Keine aufdringlichen Fragen.
- Einfache Bearbeitung.
- Klare Struktur, ansprechendes Äußeres.
 - Schriftart, Schriftgröße
 - Konsistenz der Schrifteffekte (z.B. Fett immer Erklärungen).
 - Internet: Navigation, benutztes Frontend.
- Motivation schaffen.
 - Besser genau erklären, warum diese Umfrage stattfindet und welchen Vorteil der Teilnehmer von seiner Teilnahme und dem Ergebnis hat.
 - Geschenke hingegen verfälschen eventuell das Ergebnis.
- Hintergrundinformationen liefern.
 - Wer bezahlt die Studie.
 - Nennen einer Kontaktperson.
- Angabe einer geschätzten Zeit für die Beantwortung.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 24/29

Vorbereitung der Auswertung

Partitionierung der Antworten.

- Wenn weitere Feinheiten eingegrenzt werden sollen.
- Zielgruppe muss sinnvoll weiter eingegrenzt werden.

Vereinbarungen treffen zum

- Umgang mit nicht beantworteten Fragen
 - Nur Frage weglassen? (Was bei Relationen?)
 - Ganzen Bogen weglassen? (valid sample?)
- Codierung der Daten
 - 0 oder 1 statt Nein oder Ja, Schulnoten statt Skalenantworten. Wichtig für Rechnerverarbeitung.
 - Wie umgehen mit Antworten auf offene Fragen? Nacharbeitung durch erfahrene Leute.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 25/29

Auswertung

Die eigentliche Auswertung wird mit den üblichen statistischen Methoden (Mittelwerte, Varianzen etc.) durchgeführt.

- Auf diese wird hier nicht genauer eingegangen.
- Aber Achtung: Bei Umwandlung von Skalen in Zahlen ist in Verbindung mit Mittelwertbildung Vorsicht angesagt.

Kategorie	Frage 1 (green)	Frage 2 (blue)	Frage 3 (purple)
++	10	30	20
+	20	10	20
0	40	10	20
-	20	15	20
--	10	30	20

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 26/29

Zusammenfassung

Eine Umfrage ist mehr als die Summe der gestellten Fragen. Die klar formulierte Hypothese oder Forschungsfrage und die genaue Kenntnis der Zielgruppe sind für den Erfolg der Studie entscheidend.

Bei der Ausarbeitung der Fragen und deren Präsentation sollte eine neutrale, angemessene, nicht beeinflussende Form gewählt werden.

Die korrekte Auswahl der repräsentativen Stichprobe ist genau zu prüfen.

Es reicht nicht, die Antworten zu zählen, man muss auch wissen, wie viele hätte antworten können.

Die meisten Umfragen und deren Ergebnisse, die uns im täglichen Leben begegnen, sind nicht aussagekräftig.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 27/29

Kritik an der Vorlage

Kein direkter Zusammenhang zur Softwaretechnik erkennbar, obwohl die Autorinnen dies extra betonen.

Zwar Beispiele aus der Softwaretechnik, werden aber aus dem Kontext gerissen und sind damit nicht sinnvoll verwendet.

Vorlage ist keine Studie, sondern beruht auf persönlichen Beobachtungen (gesunder Menschenverstand) oder gibt nur Inhalt anderer Quellen wieder.

Nachteil von Umfragen an sich: Sie geben ein subjektives Bild der Realität wieder, das nur schwer messbar und übertragbar ist.

© Martin Spickermann (spickerm@inf.fu-berlin.de) • 31.03.04 • 28/29

