

The melting pot - practical experience of scanning paper forms

By Marianne Troelsen & Lars Pedersen, Statistics Denmark

1. Abstract

History has proven paper as the number one carrier of questionnaires with many hard-to-beat features. Responding to a questionnaire is in many cases based on the good will of a respondent. Even with the best technological solution, some respondents will always prefer paper to any other solution. And pressed to any other solution will affect the overall response rate.

On the other hand, just because a proportion of our respondents prefer paper to electronic carriers – the NSI's do not have to bare the burden of paper. With that in mind we set out to define a system that makes the need for paper redundant. The basis is a result of a scanning process, where an image and a data file is created for each scanned questionnaire. A prototype was created, that visualized the key feature, which is, the ability to combine image and data in one screen as seen in screenshot 1.

Beskæftigede ved bygge- og anlægsvirksomhed 12 15-41-36-89

Murerarbejder og entreprenører

Hvilke personer skal medtages i opgørelsen?
 I opgørelsen medtages personer, der er ansat til virksomheden i løst eller fast arbejde i forbindelse med bygge- og anlægsvirksomhed.

Følgende personer skal ikke medtages:
 - personer for udelukkende arbejde i udførelse af bygge- og anlægsvirksomhed i administration og drift af andre aktiviteter end bygge- og anlægsvirksomhed
 - personer på barsels-, uddannelses-, selskabs- eller statsforbehold

Bygninger omfatter: Skoler, bygninger til boligformål, administration og kontorbygninger, hoteller, forretningsbygninger, industribygninger, institutionsbygninger, sportsanlæg o. s. v.

Anlæg omfatter: veje, anløb, broer, havne, alle- og gasledninger, fælledanlæg, vandledninger, vaskeri og carportanlæg, sportstøtter o. s. v.

Skemaet benyttes til indberetning af murerforretninger samt entreprenører, herunder vedrørende bygge- og anlægsvirksomhed, undtagen bygge- og anlægsvirksomhed.

Antal arbejdere og medarbejdere mædt, der d. 13. feb. 2002 udførelse eller hovedsageligt er beskæftiget med:	Medarbejdende mædt	Faglært	Ikke-faglært	Lærings-/AUB-elev	Beskæftigede d. 13. feb. 2002
1. Bygningsarbejde: opbygning og tilbygning, herunder grundarbejde, jord- og kloakarbejde, plæning og vanding i forbindelse med opførelsen af bygninger			9		9
2. Bygningsarbejder: reparation og vedligeholdelse, herunder også vedligeholdelse samt arbejde i eksisterende bygninger og forbindelse med ombygninger			2		2
3. Anlægsarbejde: opstilling og indretning af anlæg, samt reparation og vedligeholdelse heraf, herunder fx veje, jorde og kloakarbejde, portarbejde af enhver art (se dog punkt 1), kazer, bro-, havnearbejde o. s. v.			19		19
4. Andet arbejde, der ikke har direkte forbindelse med opførelse og reparation af bygninger, herunder arbejde med anlæg, herunder udførelse af bygge- og anlægsvirksomhed, herunder murerarbejde, støbe- og støbebetonarbejde o. s. v.			3		3
5. Ikke på arbejde som følge af dårligt vejr			3		3
6. Ikke på arbejde som følge af ferie, sygdom, strejke, undervisning o. l.			3		3
7. Antal arbejdere og medarbejdere mædt i alt (1+2+3+4+5+6)			33		33

Øvrige ansatte

	Mænd	Kvinder
8. Andre vigtige medarbejdere eller normalt ikke medtaget i forbindelse med opførelse, herunder også evt. angivelser, hvis der er beskæftiget i virksomheden	1	1

Journal: 15413689
 Tælling: 13-02-2002
 Nybygning: 9
 Reparation: 2
 Anlægsarb: 19
 Andre Arb: 3
 Vejr: 3
 Ej Arb: 3
 AntArb: 33
 AuxAntArb: 36
 AnsVndeh: 1
 Funktions: 1
 AndreAns: 1
 DevigAns: 2
 AuxAntDev: 2
 AnsAlt: 35
 AuxAntAlt: 38

Also the active entry point is marked simultaneously in both image and the editable data field. To ease the error-handling process symbols for warnings and absolute errors are used. In the prototype the search facilities are very limited.

2. The prototype – step by step

The prototype was written entirely in Blaise:

- 1) Define a data-model for the questionnaire, including the field-definition and rules-section (where the error-checking is defined)
- 2) Make a definition of the questionnaire in the ZoneDef program. This is where the link between data and the precise zone in the questionnaire is done.
- 3) Make a manipula-program, which combines the data-file and the image file. The key to the match is in this case an identification number.
- 4) Now all there is to do is running the manipula-program. This results in a database, filled with scanned data. Now it is possible to see each questionnaire on screen with the corresponding data in the dep-window (Screenshot 1).

In the prototype it wasn't interesting to see all the clean questionnaires, therefore some additional manipula-programs were used.

- 1) One that runs through the questionnaires checking the rules.
- 2) One that counts all questionnaires, to find clean, suspect, dirty, notchecked.
- 3) One that moves clean questionnaires to a clean database

Now, all you have left is the questionnaires, which need to be checked. It is possible to correct obvious faults, ie. the scan, respondents writing in the wrong columns. Then run the check-programs again, and move the (new) clean questionnaires to the clean database, thus reducing the dirty database to the questionnaires, that need direct action with the respondent.

3 First version

With the key functionality in place a system was planned to handle all tasks usually needed to handle paper questionnaires. The tasks in question are error handling and normal everyday retrieval of a specific questionnaire.

Some of the requirements for the solution are:

- The user interface must be menu driven. It shall be possible to get a status on a chosen dataset as to number of records, how many errors and a count on error types.
- Multi user system.
- Possibility to add (also without an image file) and delete a questionnaire.
- Filter. Example: Search for records with specific error-type and possibility to auto-jump between records with the same error-type. Or retrieve all questionnaires from a postal code. Or ..
- Search. In all fields – with wildcard characters.
- Easy access to relevant additional metadata in list boxes.