

DEZE HELPFIL E KWICDEXHELP.PDF GEEFT AANWIJZINGEN VOOR HET GEBRUIK VAN KWICDEX.

---

## **Inhoudsopgave**

### **Inleiding**

([Kwic](#), [verwijzing naar acte-nummer](#), [secundaire verwijzing](#), [meerdere soorten indices](#), [nadelen](#), [extra vermeldingen](#))

### **Werkwijze**

([In hoofdlijnen](#), [voorbeeld van invoerfile](#), [nadere details html-bewerking](#), [harde spaties en zachte koppeltekens](#), [diacrieten](#))

### **Invoerformat**

([Format van het invoerbestand](#), [werking van het programma](#))

### **Kwicdex**

([Installeren](#), [commando's](#), [boodschappen](#))

### **Opties**

([Markeringen](#), [Uitvoer keuzes](#), [opstartparameter](#), [beperkingen](#))

### **Tekencodes**

([Drie manieren om 'vreemde tekens' in te voeren](#), [html-tekencodering](#))

Opm.: Deze inhoudsopgave is in de bookmarks van de reader zichtbaar.

---

**Huib J. Plankeel**  
**Leiden, 8 april 2006.**

---

## Inleiding

Het programma **Kwicdex** is een hulpprogramma om een zogenaamde KWIC-index te maken. Het is ontstaan vanuit de wens, een goede index te maken op mijn bewerking van de schepenacten van Zierikzee van omstreeks 1600. De naamgeving in die tijd is dusdanig dat een ingang/sleutelwoord op 'Janssen' volstrekt ontoreikend is.

Het acroniem **KWIC** staat voor 'keyword in context'. In een KWIC-index wordt niet één sleutelwoord ('keyword') geïndexeerd, maar een 'indexterm', die uit meerdere elementen (woorden) bestaat. Alle elementen komen daarin als sleutelwoord voor, vergezeld van de andere elementen uit die indexterm. Als voorbeeld (fictief) de indexterm: 'Janneke van Dalen wed. Karel Hoogdorp'. Deze indexterm bestaat uit zes elementen en elk element wordt een sleutelveld; de andere elementen vormen daarbij de context.

	Janneke	Janneke van Dalen wed. Karel Hoogdorp
	Janneke van	van Dalen wed. Karel Hoogdorp
	Janneke van Dalen	Dalen wed. Karel Hoogdorp
	Janneke van Dalen wed.	wed. Karel Hoogdorp
	Janneke van Dalen wed. Karel	Karel Hoogdorp
		Hoogdorp

In dit voorbeeld zijn de eerste woorden in de tweede kolom de sleutelwoorden. De andere woorden of elementen geven de context.

Twee dingen zijn direct duidelijk. Een KWIC-index is informatief en volumineus.

Daarnaast waren er een aantal andere overwegingen en wensen:

- Een verwijzing naar een paginanummer, zoals gebruikelijk bij tekstverwerkers, heeft betrekkelijk weinig zin, wanneer er zeer veel namen op een pagina staan. Beter zou een [verwijzing naar een 'acte-nummer'](#) of iets dergelijks zijn. Bij de normale tekstverwerkers is zo'n optie evenwel niet beschikbaar.
- Mijn Zierikzeese brondocument is voor het publiek beschikbaar op microfiche, en de microfiche-beeldnummers zijn dus een tweede verwijzing. Het zou handig zijn als deze [secundaire verwijzing](#) eveneens in de index zou staan, zodat de gebruiker van de index niet een tweede lijst nodig heeft, om de juiste microfiche te vinden.
- Het zou handig zijn, wanneer in één actie [verschillende soorten indices](#) gemaakt zouden kunnen worden, zoals op personen, beroepen, plaatsen enz.
- Het programma moest verder niet te moeilijk zijn; het maken van een index is al moeilijk genoeg.

Het belangrijkste [nadeel](#) van deze methode van indexeren is, dat er geen koppeling is tussen brondocument (dus de oorspronkelijke tekst) en het indexdocument. Een verandering in het brondocument wordt niet automatisch in het indexdocument verwerkt, het zijn immers twee verschillende documenten.

Om dat nadeel (gedeeltelijk) op te vangen, zou men de kleurmarkeringen ook in het brondocument kunnen aanbrengen. Weliswaar is dan nog steeds de conversie via html naar kwicdex-invoerfile nodig, maar dat valt in hoge mate te automatiseren. Het brondocument ziet er dan behoorlijk gekleurd uit, maar vaak is het mogelijk bij afdrucken de optie alles zwart afdrucken te kiezen. En men zou ook ten behoeve van verspreiding een 'schone' kopie kunnen maken, waarin alle kleuren op automatisch of zwart gezet zijn.

## Aanwijzingen voor het gebruik van Kwicdex.

Soms is het nodig in het indexdocument [extra vermeldingen](#) te maken. Wanneer men de naam 'Joris Lenerts schoemaker' heeft (om hem te onderscheiden van 'Joris Lenerts visser'), dan zou men ook een aparte ingang op het beroep 'schoenmaker' willen hebben. Omdat een woord maar één kleur kan hebben, vereist dat dus een extra vermelding. Deze extra vermeldingen maken het brondocument dan 'onduidelijk', tenzij men de extra vermelding als [notitie](#) invoegt. [Notities worden namelijk ook naar html uitgevoerd](#). Alleen kan men notities niet kleuren, maar men kan natuurlijk wel de kleurindex handmatig aan de notitie toevoegen, dus b.v. '{\$2 schoemaker}'. Het is dus een beetje afhankelijk van het soort document en het soort index, dat men wil hebben, wat de handigste methode is.

## Werkwijze

### In hoofdlijnen:

Na wat experimenten bleek de handigste manier om een index te maken via html-uitvoer te lopen, als volgt:

1. In een kopie van het te indexeren document worden (in de tekstverwerker) alle 'indextermen' met een kleur gemarkeerd. Daarbij kan per soort index een andere kleur gekozen worden, b.v. personen: rood, beroepen: blauw, enz. Het is daarbij handig aan de cijfers 1 t/m 9 van het numerieke toetsenbord een macro te koppelen, die de geselecteerde indexterm een bepaalde kleur geeft.
2. Men controleert de tekst nu op de door Kwicdex gebruikte stuurtekens.
3. Daarna wordt dit document als html-document opgeslagen, waarna de html-broncode verder wordt bewerkt. Het is daarbij in elk geval nodig om de extensie van [.html](#) te veranderen in [.txt](#), en om het document expliciet als [.txt document](#) in de tekstverwerker te openen. Dus niet via dubbel klikken op de naam; dat werkt hier meestal niet goed.
4. Via zoek/vervang wordt het document geschikt gemaakt voor invoer in Kwicdex. Dit lijkt complex maar is het niet. Kwicdex moet een aantal dingen weten om de index te kunnen maken, b.v. waar begint een 'record', waar begint een 'indexterm', enz. Daarvoor kijkt Kwicdex naar bepaalde codes in het document. Zo is '#' standaard het teken voor het begin van een record, en '|' het standaard teken voor het eind. De indexterm begint met '{\$' met achter het dollarteken een code voor de kleur. Html-code is gestandaardiseerd en het is dus betrekkelijk eenvoudig om b.v. alle '<COLOR #008000' codes te vervangen door '{\$1'. Evenzo voor de andere markeringen. De rest van de codes kan tenslotte gewoon verwijderd worden. Zie verder onder [Nadere details](#).
5. Na de zoek/vervang actie heeft men een tekstdocument dat er ongeveer zo uit ziet:

```
#f2v/3, 5 dec 1595, {$1 mr Corn. Marinussen hopman} verlijdt {$1 Corstiaen Janssen}
£32 over de koop van een {$2 overdekte schuyte}, ... {B1-003}|
#f2v/4, [5 dec 1595], [niet gepasseerd], {$1 mr Pieter Ælas} transporteert aan {$1
Lauris Lievenssen} twee schepenen schuldbrieven ... {B1-003}|
```

Bij '#' begint een record, en direct daarna, tot de komma, staat het eerste verwijzingsnummer; tussen '{' en '}' staan de indextermen, waarbij '\$1' en '\$2' de indexsoort aangeeft; aan het eind, eveneens tussen '{' en '}' staat het secundaire nummer, en tenslotte wordt het record afgesloten met de rechte streep '|'.  
Men ziet in dit document gemakkelijk of er nog een html-codering is blijven staan, of andere 'vreemde' tekens.

## Aanwijzingen voor het gebruik van Kwicdex.

6. Dit document wordt in Kwicdex geopend, waarna Kwicdex een indexfile maakt. Daarbij zijn nog een aantal instellingen mogelijk. Bovendien kan men in Kwicdex ook andere standaardcodes instellen. (Dat zal normaal niet nodig zijn, maar wanneer men een tekst heeft met overal '\$'-tekens, dan is het soms handiger om een andere standaard te kiezen, i.p.v. alle '\$'-tekens te veranderen.).

7. De door Kwicdex gemaakte indexfile ziet er (voor de twee voorbeeldrecords) als volgt uit:

N00002a	1		mr Corn. Marinussen hopman	f2v/3	B1-003
N00002b	1	mr Corn. Marinussen	hopman	f2v/3	B1-003
N00002c	1	mr Corn.	Marinussen hopman	f2v/3	B1-003
N00002d	1	mr	Corn. Marinussen hopman	f2v/3	B1-003
N00003a	1		Corstiaen Janssen	f2v/3	B1-003
N00003b	1	Corstiaen	Janssen	f2v/3	B1-003
N00004a	2		overdeekte schuyte	f2v/3	B1-003
N00004b	2	overdeekte	schuyte	f2v/3	B1-003
N00005a	1		mr Pieter Ælas	f2v/4	B1-003
N00005b	1	mr Pieter	Ælas	f2v/4	B1-003
N00005c	1	mr	Pieter Ælas	f2v/4	B1-003
N00006a	1		Lauris Lievenssen	f2v/4	B1-003
N00006b	1	Lauris	Lievenssen	f2v/4	B1-003

Na sortering op kolom 2, 4, en 1 hebben we een index, eerst op personen (index 1) en daarna op zaak (index 2):

N00005b	1	mr Pieter	Ælas	f2v/4	B1-003
N00002d	1	mr	Corn. Marinussen hopman	f2v/3	B1-003
N00003a	1		Corstiaen Janssen	f2v/3	B1-003
N00002b	1	mr Corn. Marinussen	hopman	f2v/3	B1-003
N00003b	1	Corstiaen	Janssen	f2v/3	B1-003
N00006a	1		Lauris Lievenssen	f2v/4	B1-003
N00006b	1	Lauris	Lievenssen	f2v/4	B1-003
N00002c	1	mr Corn.	Marinussen hopman	f2v/3	B1-003
N00002a	1		mr Corn. Marinussen hopman	f2v/3	B1-003
N00005a	1		mr Pieter Ælas	f2v/4	B1-003
N00005c	1	mr	Pieter Ælas	f2v/4	B1-003
N00004a	2		overdeekte schuyte	f2v/3	B1-003
N00004b	2	overdeekte	schuyte	f2v/3	B1-003

Voor alle duidelijkheid: alleen de eerste woorden in kolom 4 zijn de sleutelwoorden. Kolom 3 geeft slechts de context weer, **en het heeft dus geen enkele zin op deze kolom te sorteren.**

De sortering is verder natuurlijk afhankelijk van eventuele sorteeropties, als b.v. wel/geen onderscheid tussen hoofdletters e.d.

## Aanwijzingen voor het gebruik van Kwicdex.

We zien hier op de vorige bladzij een tabel met 6 kolommen:

- kolom 1: een volgnummer, te gebruiken voor sortering e.d.
- kolom 2: de indexsoort
- kolom 3: de voorliggende context (deze kolom is rechtsuitgelijnd)
- kolom 4: het **sleutelwoord**, gevolgd door de achterliggende context
- kolom 5: de primaire verwijzing (folionummer + volgnummer)
- kolom 6: de secundaire verwijzing (microfiche beeldnummer).

**Opm.:** Het volgnummer bestaat uit een 'N', een volgnummer (aanvankelijk 5 cijfers) en een letter, waarbij 'a' de vermelding op het eerste woord aangeeft, en 'b' die op het laatste woord. (Gebruikte letters: abcdefghikmnpqrstuvw). Na N99999 gaat de telling verder met N100000.

Met deze tabel/spreadsheet kan men dan verder aan de gang, om b.v. een aparte persoons-index te maken. De volgnummers en het soortnummer zou men daarin weg kunnen laten. Men zou de records ook in een database kunnen inlezen, om daaraan later de index van een tweede document toe te voegen. Er is dan alleen een extra kolom nodig om de documenten te kunnen onderscheiden.

Ook kan men na sortering van kolom 4 gemakkelijk irrelevante ingangen/sleutelwoorden (zoals b.v. 'de' of 'aan') verwijderen. (Dat moet men natuurlijk alleen daar doen.)

### Nadere details betreffende de bewerking van het html-document:

Om het document te bewerken is met name van belang de optie '**reguliere expressie**' (regular expression) bij zoek/vervang, die men bij sommige tekstverwerkers (waaronder Word en Open Office) vindt. Deze optie biedt zeer bijzondere en geavanceerde mogelijkheden, die evenwel weinig bekend zijn. Zo kun je met '<\*>' vrijwel alle html-codes vinden (aangezien deze standaard tussen '<' en '>' staan) en in één klap met 'alles vervangen' vervangen door 'niets', en dus verwijderen. Alleen de 'meta-tags' verwijderen, gaat simpel met zoek (reguliere expressie aangevinkt): '<META\*>' en vervang alles door 'niets'. Het sterretje in de reguliere expressie is dus een joker- of wildcard-teken.

Nu moet men niet meteen beginnen om alle html-codes te verwijderen, want dan worden ook de kleurcoderingen e.d. verwijderd. Ook moet men ervoor zorgen, dat men de html-codes kan gebruiken om begin en einde van de records te markeren.

In html eindigt een paragraaf (d.w.z. een stuk tekst zonder return erin) op '</p>'. We kunnen dit alvast (met gewoon zoek/vervang) vervangen door het voor Kwicdex benodigde **einde-record-teken** '|'. En het is handig daar meteen een nieuwe-regel-teken aan toe te voegen. (Dat geeft zowel een beter visueel overzicht, maar voorkomt ook, dat later per ongeluk twee zinnen aan elkaar geplakt worden, zonder spatie er tussen). Het hangt van de tekstverwerker af, of dat bij de gewone opties staat, of dat dat via reguliere expressies moet.

**N.B.** Dit teken '|' mag niet elders in de tekst voorkomen, dus het is handig om voor deze actie even te zoeken of dit teken voorkomt, en zo ja, maatregelen te nemen, b.v. door het te verwijderen, of door een ander teken te vervangen. Als dat een probleem zou zijn, dan kan in Kwicdex ook een ander einde-record-teken ingesteld worden.

De volgende stap is het **recordbegin**. Kwicdex bepaalt het pagina/recordnummer op basis van de tekst tussen **recordbeginteken** '#' en de eerstvolgende **komma** of **spatie**. Het hangt van het document af, wat hier de beste oplossing is. Bij een structuur als de volgende: '<P CLASS="western" STYLE="margin-bottom: 0cm">RAZE 3851b, f2/1, ' begint elk record met 'RAZE 3851b, ' en wanneer men dat vervangt door '#' dan heeft men als recordbegin '#f2/1, '.

## Aanwijzingen voor het gebruik van Kwidx.

De html-code '<P ...>' kan dan, zoals hiervoor beschreven met (reguliere expressie) zoek '<P\*>' en vervang door 'niets', verwijderd worden. Als de documenttekst evenwel geen duidelijk begin heeft, dan moet men wellicht ook deze '<P...>' code gebruiken om de begin recordmarkering aan te brengen, b.v. via (regulier) zoek '<P\*f' en vervang door '#f'. Dit is natuurlijk een veel riskantere operatie om ongecontroleerd te doen, want stel dat ergens de 'f' voor het folionummer vergeten is!

Dan de [markeringen voor de indextermen](#). Ook hier is de html-code standaard, al verschilt de uitvoer wel per tekstverwerker: in OpenOffice staat b.v. '<FONT COLOR="#000080">Matheus Pieterssen</FONT>'. In Word uitvoer staat het weer anders, maar herkenbaar is het altijd. Het getal achter het 'hekje' geeft de kleur aan, en bij gebruik van meerdere kleuren heb je hier dus andere waarden staan. Voor Kwidx moet de indexterm beginnen met '{ \$1 ' waarbij de '1' de kleur aangeeft; i.p.v. '1' mag overigens alles staan, ook b.v. K1 of persoon. Let evenwel op de [spatie](#) erachter. Ook hier weer een simpele (niet reguliere) zoek/vervang actie: zoek '<FONT COLOR="#000080">' en vervang door '{ \$1 '. Doe evenzo met de andere kleuren, maar met natuurlijk een ander nummer dan '1'. Tot slot moet '</FONT>' aan het eind vervangen worden door het sluithaakje '}'.

De tekst begint er nu al aardig schoon uit te zien. Het is nu zinnig naar de overblijvende codes te kijken: B.v. het voetnootnummer: '<A CLASS="sdfootnoteanc" NAME="sdfootnote1anc" HREF="#sdfootnote1sym"><SUP>1</SUP></A>'. In het geval, dat het voetnootnummer in de indexterm voorkomt, moet niet alleen de code, maar ook het nummer (hier '1') weggehaald worden, anders komt dat in de index terecht. De (reguliere expressie) zoek/vervang actie wordt dan: zoek '<A\*/A>' en vervang door 'niets'.

Controleer nu de tekst om te zien, of er nog 'rare' dingen in zitten. Kijk vooral of een record niet in twee stukken (paragrafen) is gesplitst. Dat kan gebeuren, wanneer men in het origineel een 'enter' gegeven is (want dan wordt een nieuwe paragraaf gestart) i.p.v. een 'shift-enter' (daarbij blijft men in dezelfde paragraaf). Tot slot kunnen alle overige html-codes verwijderd worden.

De volgende 'zoek/vervang' actie betreft de [html-tekencoderingen](#). Bijzondere tekens als 'letters met accenten', 'groter' en 'kleiner' teken, 'pond'-teken enz. worden in html weer gegeven door een '&', gevolgd door wat letters, en afgesloten met een puntkomma.

In mijn document b.v. stond de secundaire verwijzing tussen '<' en '>'. Bij de omzetting naar html is dat veranderd in '&lt;fiche B1-003&gt;'. Omdat ik voor Kwidx de accolades als markering gebruikt moet ik '&lt;fiche ' vervangen door '{' en '&gt;' door '}'. Het woord fiche mag in de index wel vervallen. Dat scheelt ruimte. Let er dus op, dat er niet per ongeluk een '}' wordt neergezet op een plek waar dat niet moet, doordat in de originele tekst b.v. een '>' stond. Een goede controle is, om bij te houden hoeveel tekens vervangen werden. Bij tekens, die in paren voorkomen, zouden beide aantallen gelijk moeten zijn.

Twee bijzondere tekens zijn de '[harde spatie](#)' (in html '&nbsp;') en het '[zachte koppelstreepje](#)' (soft hyphen; in html '&shy;').

Een '[harde spatie](#)' ziet eruit als een spatie maar wordt beschouwd als een gewoon teken. Een regel wordt daar niet afgebroken.

Een '[zacht koppelstreepje](#)' is een onzichtbaar teken, dat men in een lang woord kan plaatsen, om ervoor te zorgen, dat het aan het eind van een regel op die plek wordt afgebroken.

Kwidx kan gebruik maken van het 'zachte koppelstreepje', om, net als bij (zachte) spaties, een term te splitsen. Dat is bruikbaar om b.v. in een beroepenindex niet alleen een ingang op 'huistimmerman' te hebben, maar ook op 'timmerman'. Achter 'huis' voegt men dan een (onzichtbaar) 'zacht koppelstreepje' toe.

## Aanwijzingen voor het gebruik van Kwicdex.

Evenzo om bij een naam als 'van Bergen-Henegouwen' ook een ingang op 'Henegouwen' te krijgen. Het 'zachte koppelstreepje' voegt men in dat geval achter het 'harde koppelstreepje' toe. Op een 'harde spatie' wordt een indexterm niet gesplitst, dus als dat niet de bedoeling is dan moet de 'harde spatie' op die plek vervangen worden door een 'zachte spatie'.

Zoeken/vervangen gaat als volgt:

zoek '&nbsp;';,            vervang door linker alt toets plus 0160 (numerieke toetsenbord)

zoek '&shy;';,        vervang door linker alt toets plus 0173 (numerieke toetsenbord).

Een andere grote groep tekens zijn de '[diacrieten](#)', dus de tekens met een accent, als á, ë, e.d. Hier moet men bedenken wat in de index handig is. Een 'ë' is immers een ander teken dan 'e' en sorteert daarom (afhankelijk van het programma) anders. Wanneer men in de index geen onderscheid wil tussen 'Daniel' en 'Daniël' dan moet men '&euml;'; vervangen door 'e'. Er zijn tabellen voor deze zogenaamde '[html character entities](#)' maar men kan natuurlijk ook kijken wat er in de oorspronkelijke tekst stond.

Na al deze veranderingen, is het handig nog even te controleren of er niet één vergeten is, door (via reguliere expressie) te zoeken op '&\*';.

Na al deze zoek/vervang handelingen moet men een bestand hebben gekregen, dat geschikt is voor invoer in Kwicdex, en dat er ongeveer [zo](#) uitziet.

**Opm.:** Veel van deze zoek/vervang acties zijn, bij gebruik van een bepaalde tekstverwerker, redelijk standaard, zeker wanneer men op een standaard manier werkt. Wanneer men via macro's de kleuren aan het numerieke toetsenbord heeft toegewezen, zal men in elk bestand, dat als html wordt uitgevoerd, dezelfde html-kleurcodering vinden. Men kan dus ook een simpele zoek/vervang macro maken, die negen kleurcodes plus een slotcode automatisch omzet. Hetzelfde geldt voor de html-tekens.

## Invoerformat

### Format van het invoerbestand

**Kwicdex** verwacht een invoerfile, die op een bepaalde manier gemarkeerd is:

```
#f2v/3, 5 dec 1595, {$1 mr Corn. Marinussen hopman} verlijdt {$1 Corstiaen Janssen} £32  
over de koop van een {$2 overdeekte schuyte}, ... {B1-003}|
```

```
#f2v/4, [5 dec 1595], [niet gepasseerd], {$1 mr Pieter Ælas} transporteert aan {$1 Lauris  
Lievenssen} twee schepenen schuldbrieven ... {B1-003}|
```

Bij '#' begint een record, en direct daarna, tot de komma, staat het eerste verwijzingsnummer; tussen '{' en '}' staan de indextermen, waarbij '\$1' en '\$2' de indexsoort aangeeft; aan het eind, eveneens tussen '{' en '}' staat het secundaire nummer, en tenslotte wordt het record afgesloten met de rechte streep '|'.

De vijf (of zes) standaard markeringen mogen alleen op de voorgeschreven plaats voorkomen. Er zijn in het programma enkele [alternatieven](#) te kiezen.

## Werking van het programma

In detail werkt het programma zo:

**Kwicdex** leest het bestand, totdat het een **recordbeginteken** (standaard '#') tegenkomt. Alles wat daarna komt, tot het **einderecordteken** (standaard '|') wordt in een **buffer** (van maximaal 30000 tekens) gezet. Een 'nieuwe regel'-teken wordt in een spatie omgezet, en als er twee spaties staan, wordt er één verwijderd.

**Opm.:** Als het einderecordteken ontbreekt, wordt het record overgeslagen!

Het programma ziet immers weer een volgend recordbeginteken.

Daarna wordt de inhoud van de buffer verwerkt:

Alles na het **recordbeginteken**, tot de eerste **komma** of **spatie**, wordt beschouwd als **primair verwijfsnummer**. Dat kan zowel een folionummer zijn (b.v. f12v) of een nummer (n001), maar ook b.v. een datum (12-10-1754) of een paragraaf nummer (§1.3). Alles kan, mits er geen (zachte) spatie of komma in staat. Omdat een spreadsheet soms problemen heeft met de correcte interpretatie van getallen, of denkt dat '=' of '@' naar een cel of een formule verwijst, is het handig ervoor te zorgen dat het verwijfsnummer altijd begint met een letter. Dan ziet de spreadsheet het altijd als tekst. Dat geeft de minste problemen.

Vervolgens wordt (maar alleen als de Uitvoer keuze '**Secundair nummer aanwezig**' is aangevinkt), *vanaf het einde van de buffer terug* gezocht naar een **Begin indexomsluiting teken** (standaard '{'). Als dat gevonden wordt, dan wordt vandaar af weer naar het eind gegaan tot de eerste **komma**, en alles daartussen wordt beschouwd als **secundair verwijfsnummer**. Wanneer evenwel een Indexmarkering gevonden wordt, dan volgt een waarschuwing, en bestaat de mogelijkheid het programma af te breken. Net als bij het primaire nummer zijn er (buiten de standaardcoderingstekens) geen beperkingen. Als tussen de haakjes *alleen een spatie* staat, dan wordt het vorige verwijfsnummer gebruikt.

Daarna zoekt het programma vanaf het begin van de buffer naar de eerste **Indexmarkering** (standaard '\$'). Alles wat daarachter staat, tot de **eerste spatie** (**hier dus geen komma!**), wordt beschouwd als **indexsoort**. Dat kan zowel een nummer zijn, bv. K1, maar ook een woord als 'persoon', 'zaak', 'beroep'. Vanaf deze spatie tot het **Einde indexomsluiting teken** (standaard '}') wordt beschouwd als **indexterm**. Een eventuele spatie voor het haakje wordt verwijderd.

**Opm.:** Het **Begin indexomsluiting teken** (standaard '{') dient vooral ter controle.

Daarna worden op basis van de '**Uitvoer keuzes**' per term vermeldingen gemaakt en naar het uitvoerbestand weggeschreven. Vervolgens worden de volgende **Indexmarkeringen** in de buffer verwerkt. Daarna wordt het volgende record in de buffer gezet en verwerkt, net zolang tot alle records verwerkt zijn. Het resultaat wordt standaard naar een bestand geschreven met de zelfde naam als het invoerbestand, maar met de extensie **.csv**. Als dat bestand al bestaat, dan verschijnt een bestandsmenu en kan een andere naam gekozen worden, of men kan het bestand overschrijven.

## Kwicdex

### Installatie

Het **programma Kwicdex** is een 32-bit Windows toepassing, geschreven in Delphi. Het bestaat uit slechts één kleine '.exe'-file, en behoeft niet geïnstalleerd te worden. Het brengt geen veranderingen aan in de registry noch in enig ander systeem bestand. Nadat Kwicdex in een map is gezet, kan men een koppeling maken naar het bureaublad, en het vandaar starten. Indien gewenst kan men een '**opstart-parameter**' mee geven (via eigenschappen van de icoon) om alvast bepaalde **opties** te selecteren. Ook de helpfile **KwicdexHelp.pdf** hoort in die map.

## Commando's

Naast de onder '[Opties](#)' beschreven 'keuze blokken' zijn na opstarten twee of drie knoppen zichtbaar. '[Open Invoerfile](#)' opent een standaard Windows file-menu, waarin men het te verwerken bestand kan selecteren. Na selectie gaat het programma direct van start, tenzij de 'standaard' uitvoerfile al bestaat. Standaard wordt de uitvoer namelijk geschreven naar een bestand, in dezelfde map en met dezelfde naam als de invoerfile, maar met de extensie '.csv'. Deze extensie duidt een 'komma' gescheiden bestand aan. Als zo'n bestand reeds bestaat, dan verschijnt een tweede Windows file-menu (met titel: [Hernoem of overschrijf uitvoerfile](#)), waarin men zonnodig een andere bestandsnaam kan kiezen. De 'standaard uitvoerfile naam is reeds aangegeven, en na het drukken op Save wordt gevraagd of het bestand overschreven mag worden. Bij 'Cancel' wordt het programma afgebroken. '[Stop](#)' spreekt voor zich.

Terwijl het programma bezig is, loopt er op het scherm een tellertje, zodat men kan 'zien', bij welk recordnummer het programma is. Als het programma klaar is, wordt aangegeven hoeveel vermeldingen zijn uitgevoerd, en de namen van invoer- en uitvoerfile.

Wanneer de helpfile '[KwicdexHelp.pdf](#)' (d.w.z.: dit helpbestand) en het programma Kwicdex beide in dezelfde map staan, dan is er nog een derde knop '[Help](#)' aanwezig. Als men daarop klikt, wordt deze helpfile in de Adobe/Acrobat-Reader geopend (mits er onder Windows een koppeling bestaat tussen de extensie .pdf en de reader). Als de helpfile niet aanwezig is, of elders staat, dan vindt men slechts een aanwijzing op het scherm.

## Boodschappen:

- [Kies naam voor uitvoerfile.](#)  
Het programma moet een bestand openen om de gegevens naartoe weg te schrijven.
- [Er ontbreekt een haakje voor het secundair nummer, afbreken?](#)
- [Onverwacht indexvermelding gevonden i.p.v. secundair nummer, afbreken?](#)
- [Er ontbreekt een haakje achter een markering, afbreken?](#)  
In deze gevallen is er nu de mogelijkheid het programma af te breken, om b.v. de optie 'Secundair nummer' goed in te stellen. In andere gevallen kan men beter het programma gewoon laten doorlopen, en registreren waar fouten staan. De uitvoerfile heeft hiervan verder geen last.
- [Geen indexvermeldingen aangetroffen. Programma is nu gestopt; de uitvoerfile is leeg.](#)  
Dit kan eigenlijk alleen gebeuren, wanneer men een 'verkeerd' bestand opent, zonder coderingen, e.d. Omdat de uitvoerfile aangemaakt wordt, voordat het programma de buffer gaat vullen en bewerken, kan in dat geval een leeg bestand aangemaakt zijn.

## Opties, mogelijkheden en beperkingen

Het startscherm van Kwicdex toont naast de twee of drie hiervoor beschreven knoppen ook twee panels, met resp. '[Uitvoer keuzes](#)' en '[Markeringen](#)'.

## Markeringen

In 'Markeringen' zijn de tekens zichtbaar, die Kwicdex als stuurtekens hanteert. Wanneer de bestaande tekens niet voldoen, dan zijn hier eventueel enkele andere te selecteren. Deze instellingen worden niet bewaard. Wel kan men ze vooraf via een [Opstartparameter](#) instellen.

## Aanwijzingen voor het gebruik van Kwicdex.

Om nog even aan te geven hoe je enkele van deze speciale tekens maakt:

	rechter 'alt-toets'	linker 'alt-toets'
omgekeerd uitroepteken 'i''	'1'	numeriek 0161 (of 173)
het 'not teken' '¬'	'\'	numeriek 0172 (of 170)
omgekeerd vraagteken '¿'	'/'	numeriek 0191 (of 168)

## Uitvoer keuzes

Onder 'Uitvoer keuzes' kan men een aantal opties kiezen:

- **Zacht koppelstreepje (soft hyphen; in html '&shy;')** als extra marker.  
Een 'zacht koppelstreepje' is een onzichtbaar teken, dat men in een lang woord kunt plaatsen, om ervoor te zorgen, dat het aan het eind van een regel op die plek wordt afgebroken. Als deze optie aan staat, maakt Kwicdex ook gebruik van het 'zachte koppelstreepje' om een term te splitsen. Dat is bruikbaar om b.v. in een beroepenindex niet alleen een ingang op 'huistimmerman' te hebben, maar ook op 'timmerman'. Achter 'huis' voegt men dan een (onzichtbaar) 'zacht koppelstreepje' toe. Evenzo om bij een naam als 'van Bergen-Henegouwen' ook een ingang op 'Henegouwen' te krijgen. Het 'zachte koppelstreepje' voegt men in dat geval achter het 'harde koppelstreepje' toe. Zoeken/vervangen gaat in het html-bestand als volgt:  
zoek '&shy;', vervang door linker 'alt-toets' plus 0173 (numerieke toetsenbord).
- **Secundair nummer aan het eind aanwezig.**
- **Alleen elementen met hoofdletter indexeren.**  
Dit is vooral bedoeld voor indexeren van namen. Als deze optie gekozen is, dan zullen de meeste zaken niet in de index komen, dus daarvoor moet je het programma dan nog een keer draaien.  
Als **hoofdletter** beschouwt het programma de tekens 'A' t/m 'Z', de hoofdletters met diacrieten, volgens de html-4 standaard de nummers &#192 t/m &#214, en &#216 t/m &#222. Daarnaast nog de vier niet standaard hoofdletters met nummers &#138 (Š), &#140 (Œ), &#142 (Ž) en &#159 (Ÿ).
- **(alleen) index op eerste element.**  
Als dit is aangevinkt, dan wordt de vermelding uitgevoerd, waar het eerste element het sleutelwoord is. Dus van de indexterm 'mr Jan de Bakker junior' wordt alleen de vermelding op sleutelwoord 'mr' uitgevoerd.  
**Opm.:** Deze optie wordt beïnvloed door de voorgaande optie 'alleen hoofdletters ...'; als die optie aan staat, wordt de vermelding met sleutelwoord 'Jan' uitgevoerd.
- **(alleen) index op laatste element.**  
Hier geldt hetzelfde als bij de vorige optie, maar dan betrekking hebbend op het laatste woord. Ook hier weer de beïnvloeding door de 'hoofdletter'-optie.  
  
Beide opties zijn onafhankelijk van elkaar te gebruiken. Als beide zijn aangevinkt, dan komt er (grof gezegd) een index op voornaam en een index op achternaam. Natuurlijk is een en ander zeer afhankelijk van de markeringen. Wanneer men als naam gemarkeerd heeft 'mevr. Jansen-van de Berg' en de optie 'laatste element' is aangevinkt, dan wordt alleen een vermelding op 'Berg' gemaakt, maar niet op 'Jansen'.
- De laatste optie **'folionummer bewerken'** is er vooral voor mijn eigen gebruik. Het converteert een primair verwijzingsnummer als 'f12' in 'f012r', 'f12v' wordt 'f012v' maar ook 'f12/2' wordt 'f012r/2'. Dat is vooral om een extra sorteringmogelijkheid te hebben, al is het niet helemaal perfect, want bij twee folio's met gelijk nummer hebt, aangeduid door b.v. f12 en f12A, dan sorteert de tweede toch voor de eerste.

## Opstartparameter

Het is mogelijk het programma bij opstarten een (Command line) parameter mee te geven (dus bij het icoontje, onder eigenschappen, achter de bestandsnaam toe te voegen). Deze parameter moet uit 10 tekens bestaan, voor elke optie één. Voor de eerste zes tekens geldt dat een 'a' de optie 'aan' zet. Voor de laatste vier tekens geldt dat een nummer (van 1 tot maximaal 4) de betreffende markering instelt. Met een tweede parameter kan men het recordnummer instellen. Als hierin een fout zit, dan wordt gewoon de standaardwaarde 1 genomen. Vanzelfsprekend moet deze waarde groter zijn dan nul, als het uiteindelijk aantal records maar niet boven de 2.147.483.647 uitkomt.

## Beperkingen

Reële beperkingen kent het programma nauwelijks, op het feit na, dat het alleen standaard ANSI-tekens verwerkt, en geen UNICODE.

- De [stuurcodes](#) voor het programma mogen niet in de tekst zelf voorkomen, maar men kan desgewenst een andere set tekens kiezen. Ook via zoek/vervang kan men ervoor zorgen, dat aan deze voorwaarde wordt voldaan.
- De [invoerbuffer](#) voor een record is 30000 tekens, en dat is vele pagina's tekst.
- De [buffer voor](#) het verwerken van [spaties](#) en '[soft hyphens](#)' bestaat uit 20 posities, d.w.z. een 'indexterm' mag maximaal 20 van dergelijke tekens bevatten.
- In de [verwijzingsnummers](#) mag geen komma zitten.
- En tenslotte mag de [indexsoort](#) geen spatie bevatten.

## Tekencodes

### Drie methodes

Hoewel niet direct voor Kwicdex relevant, hier nog iets over 'internationale tekens'. Daarvoor bestaan er onder Windows drie methodes. Sommige tekens kunnen daarom op twee of drie manieren worden gemaakt. Probleem is daarbij, dat er meerdere 'code pages' in gebruik zijn om numerieke codes (met name in het bereik van 128 t/m 159) in tekens om te zetten. Twee belangrijke daarvan zijn CP 1252 (een Microsoft standaard, ook gebruikt door Delphi) en (de internationale standaard) ISO 8859-1 (en latere varianten). Hoe er vertaald wordt is dus afhankelijk van de ingestelde 'codepage'.

**Principe 1, de dode toetsen methode.** Dit vereist de installatie van een speciaal toetsenbord-programma 'VS Internationaal' (US International) en wordt ingesteld via Configuratie. (Dit zit standaard bij Windows, maar is niet altijd geïnstalleerd). Net als bij de ouderwetse schrijfmachine type je eerst een accentteken, gevolgd door een klinker; de machine zet beide tekens dan onder elkaar. Op deze manier zijn de volgende tekens te maken:

' plus aeiouy c	geeft	áéíóúý ç
' plus AEIOU C	geeft	ÁÉÍÓÚ Ç
` plus aeiou	geeft	àèìòù
` plus AEIOU	geeft	ÀÈÌÒÙ
" plus aeiouy	geeft	äëïöüÿ
" plus AEIOU	geeft	ÄËÏÖÜ
^ plus aeiou	geeft	âêîôû
^ plus AEIOU	geeft	ÂÊÎÔÛ
~ plus aonAON	geeft	ãõñÃÕÑ

## Aanwijzingen voor het gebruik van Kwicdex.

### Principe 2, de rechter alt-toets methode

Hier geeft het indrukken van de *rechter* alt-toets in combinatie met een andere toets een 'bijzonder' teken. De rechter alt-toets in combinatie met de shifttoets (sh-alt) geeft nog weer enkele nieuwe mogelijkheden. Op deze manier zijn de volgende tekens te maken:

alt met	1234567890	=	qwertyuiop[]	asd l;' z c nm, /\
geeft	ı²³¼½¾‘ ’¥x		ääé®püúíóö«»	áßð øŋ´ æ © ñµç ¿¬
sh-alt met	1 4	=	qwe tyuiop	asd l;' z c n , \
geeft	´ £	÷	ÄÅÉ ÞÜÚÍÓÖ	ÁŞÐ Ø°´´ Æ Ç Ñ Ç

### Principe 3, de 'numerieke code' methode

Hier geeft het indrukken van de *linker* alt-toets en tegelijk intypen van een viercijferige (ANSI)-code of een driecijferige (IBM)-code op het numerieke toetsenbord het 'vreemde' teken: alt 0228 en alt 132 geven beide ä. Enkele tekens, die niet op een meer simpele manier te realiseren zijn in Windows (Win 1252 codepage), zijn de volgende:

æ (alt 0154)	œ (alt 0140)	š (alt 0154)	š (alt 0138)
f (alt 0131, alt 159)		™ (alt 0153)	† (alt 0134)
± (alt 0177, alt 241)		‰ (alt 0137)	

Maar onder andere andere code pages kunnen deze nummers weer anders weergegeven worden. Overigens is het soms voor de minder gebruikelijke tekens flink zoeken, óf en zo ja, hoe ze dan weergegeven moeten worden.

### De html-tekencodes.

De speciale tekens, zoals b.v. '±' kunnen in html zowel door een (ANSI)-nummer (&#241;) als door omschrijving (in dit geval '&plusmn;') aangegeven worden. De waarden van 130 t/m 159 zijn in html niet officieel vastgelegd.

Een kort overzicht van de meer belangrijke tekens:

numeriek	naam	teken	numeriek	naam	teken
&#034;	&quot;	"	&#189;	&frac12;	½
&#038;	&amp;	&	&#190;	&frac34;	¾
&#060;	&lt;	<	&#191;	&iquest;	¿
&#062;	&gt;	>	&#198;	&AElig;	Æ
&#131;		f (niet standaard)	&#199;	&Ccedil;	Ç
&#140;		œ (niet standaard)	&#201;	&Eacute;	É
&#156;		æ (niet standaard)	&#215;	&timest;	×
&#160;	&nbsp;	'harde spatie'	&#220;	&Uuml;	Ü
&#161;	&iexcl;	¡ 'omgekeerd !'	&#223;	&szlig;	ß
&#163;	&pound;	£	&#224;	&agrave;	à
&#167;	&sect;	§	&#225;	&aacute;	á
&#169;	&copy;	©	&#230;	&aelig;	æ
&#172;	&not;	¬ 'not teken'	&#231;	&ccedil;	ç
&#173;	&shy;	- 'soft hyphen'	&#232;	&egrave;	è
&#176;	&deg;	°	&#233;	&eacute;	é
&#177;	&plusmn;	±	&#235;	&euml;	ë
&#178;	&sup2;	²	&#239;	&iuml;	ï
&#179;	&sup3;	³	&#247;	&divide;	÷
&#182;	&para;	¶	&#252;	&uuml;	ü
&#185;	&sup1;	¹	&#255;	&yuml;	ÿ
&#188;	&frac14;	¼			

\*\*\*\*\*