
Computergebruik: EXAMEN

1^e Bachelor Informatica
prof. dr. Peter Dawyndt
groep 1

donderdag 12-01-2017, 08:30
academiejaar 2016-2017
eerste zittijd

Opgave 1 (levende voormalige presidenten)

(10 pt)

Op 20 januari 2017 wordt Donald John Trump geïnaugureerd als vijfenveertigste president van de Verenigde Staten. Vanaf die dag zullen er tegelijkertijd vijf voormalige Amerikaanse presidenten nog in leven zijn.



In de geschiedenis van Amerika zijn er slechts vier perioden geweest waarin dat het geval was:

- 4 maart 1861 - 18 januari 1862: Martin Van Buren, John Tyler, Millard Fillmore, Franklin Pierce, James Buchanan
- 20 januari 1993 - 22 april 1994: Richard Nixon, Gerald Ford, Jimmy Carter, Ronald Reagan, George H.W. Bush
- 20 januari 2001 - 5 juni 2004: Gerald Ford, Jimmy Carter, Ronald Reagan, George H. W. Bush, Bill Clinton
- 20 januari 2017 - heden: Jimmy Carter, George H. W. Bush, Bill Clinton, George W. Bush, Barack Obama

Herbert Hoover leefde nog 11,553 dagen (31 jaar en 230 dagen) nadat hij zijn ambt had neergelegd. James Polk kon daarentegen slechts 3 maanden (103 dagen) van zijn pensioen genieten. In totaal zijn er acht presidenten geweest die nooit de status van "voormalig president" bereikten omdat ze overleden tijdens hun ambtstermijn: William H. Harrison (longontsteking), Zachary Taylor (buikgriep), Abraham Lincoln (vermoord), James A. Garfield (vermoord), William McKinley (vermoord), Warren G. Harding (hersensbloeding), Franklin D. Roosevelt (hersensbloeding) en John F. Kennedy (vermoord).

Opgave

Gegeven is een tekstbestand `presidenten.txt` dat informatie bevat over de levensduur en de ambtstermijn van de presidenten van de Verenigde Staten. Elke regel van dit bestand bevat informatie over één president. Deze informatie bestaat uit de volgende velden, die van elkaar worden gescheiden door tabs: *i*) volgnummer, *ii*) naam president *iii*) geboortedatum, *iv*) start ambtstermijn, *v*) leeftijd bij start ambtstermijn, *vi*) einde ambtstermijn, *vii*) leeftijd bij einde ambtstermijn, *viii*) levensduur na einde ambtstermijn, *ix*) datum van overlijden en *x*) totale levensduur. De informatievelden bevatten zelf geen tabs of puntkomma's. De eerste regel van het bestand is een commentaarregel.

De vier datumvelden bevatten een datum in het formaat *Mon d, yyyy* waarbij de maand (*Mon*) bestaat uit drie letters waarvan de eerste een hoofdletter is, gevolgd door twee kleine letters. Dagen (*d*) worden altijd aangeduid met één of twee cijfers en jaartallen (*yyyy*) met vier cijfers. Velden

met datums van gebeurtenissen die nog niet hebben plaatsgevonden zijn leeg. De vier velden die een tijdsperiode aanduiden worden zowel uitgedrukt in een aantal jaren en dagen (waarvan het aantal jaren optioneel is) en een aantal dagen (tussen ronde haakjes, waarbij de cijfers met komma's gegroepeerd worden per duizendtallen). Tijdsperiodes die nog niet afgesloten zijn, kunnen leeg zijn of gerekend worden tot op de dag van vandaag.

Gevraagd wordt om — gebruik makend van de teksteditors `vi` of `vim` — een reeks commando's op te stellen die achtereenvolgens de volgende opdrachten uitvoeren. Probeer voor elke opdracht zo weinig mogelijk commando's te gebruiken en zorg er voor dat elk van deze commando's bestaat uit zo weinig mogelijk tekens. Alle opdrachten moeten na elkaar uitgevoerd worden. Geen enkele van de commando's mag de commentaarregel bovenaan het bestand wijzigen. Ter controle kan je gebruik maken van de meegeleverde bestanden `presidenten.i.txt`, die telkens de inhoud van het bestand bevatten nadat de `i`-de opdracht werd uitgevoerd.

1. Vervang de tabs als scheidingstekens door puntkomma's (;) en verwijder alle spaties vooraan en achteraan elk veld. Zo moet

```

1 # President Birth date Presidency start date Age at start date Presidency end date Age at
2 end date Post-presidency timespan Death Date Lifespan
3 1 George Washington Feb 22, 1732 Apr 30, 1789 57 years, 67 days (20,887 days) Mar 4, 1797 65
4 years, 10 days (23,752 days) 2 years, 285 days (1,015 days) Dec 14, 1799 67 years, 295 days (24,767 days)
5 2 John Adams Oct 30, 1735 Mar 4, 1797 61 years, 125 days (22,406 days) Mar 4, 1801 65
6 years, 125 days (23,866 days) 25 years, 122 days (9,253 days) Jul 4, 1826 90 years, 247 days
7 (33,119 days)
8 3 Thomas Jefferson Apr 13, 1743 Mar 4, 1801 57 years, 325 days (21,144 days) Mar 4, 1809
9 65 years, 325 days (24,066 days) 17 years, 122 days (6,331 days) Jul 4, 1826 83 years, 82 days
10 (30,397 days)
11 4 James Madison Mar 16, 1751 Mar 4, 1809 57 years, 353 days (21,172 days) Mar 4, 1817
12 65 years, 353 days (24,094 days) 19 years, 116 days (7,056 days) Jun 28, 1836 85 years, 104 days
13 (31,150 days)
14 ...

```

bijvoorbeeld omgezet worden naar

```

1 # President Birth date Presidency start date Age at start date Presidency end date Age at
2 end date Post-presidency timespan Death Date Lifespan
3 1;George Washington;Feb 22, 1732;Apr 30, 1789;57 years, 67 days (20,887 days);Mar 4, 1797;65 years, 10 days (23,752
4 days);2 years, 285 days (1,015 days);Dec 14, 1799;67 years, 295 days (24,767 days)
5 2;John Adams;Oct 30, 1735;Mar 4, 1797;61 years, 125 days (22,406 days);Mar 4, 1801;65 years, 125 days (23,866 days);25
6 years, 122 days (9,253 days);Jul 4, 1826;90 years, 247 days (33,119 days)
7 3;Thomas Jefferson;Apr 13, 1743;Mar 4, 1801;57 years, 325 days (21,144 days);Mar 4, 1809;65 years, 325 days (24,066
8 days);17 years, 122 days (6,331 days);Jul 4, 1826;83 years, 82 days (30,397 days)
9 4;James Madison;Mar 16, 1751;Mar 4, 1809;57 years, 353 days (21,172 days);Mar 4, 1817;65 years, 353 days (24,094 days)
10 ;19 years, 116 days (7,056 days);Jun 28, 1836;85 years, 104 days (31,150 days)
11 ...

```

In het vervolg zullen we de commentaarregel weglaten uit het resultaat om plaats te besparen. Maar onthoud dat de commentaarregel door je commando's niet mag gewijzigd worden.

2. Voor presidenten die nog leven staat er geen datum in het veld met de datum van overlijden. Duidt het feit dat ze nog leven expliciet aan door de tekst `alive` in te vullen in dit veld. We krijgen dan

```

1 ...
2 41;George H. W. Bush;Jun 12, 1924;Jan 20, 1989;64 years, 222 days (23,598 days);Jan 20, 1993;68 years, 222 days (25,059
3 days);23 years, 355 days (8,757 days);alive;92 years, 211 days (33,814 days)
4 42;Bill Clinton;Aug 19, 1946;Jan 20, 1993;46 years, 154 days (16,956 days);Jan 20, 2001;54 years, 154 days (19,878 days)
5 ;15 years, 355 days (5,835 days);alive;70 years, 143 days (25,711 days)
6 43;George W. Bush;Jul 6, 1946;Jan 20, 2001;54 years, 198 days (19,922 days);Jan 20, 2009;62 years, 198 days (22,844
7 days);7 years, 355 days (2,913 days);alive;70 years, 187 days (25,755 days)
8 44;Barack Obama;Aug 4, 1961;Jan 20, 2009;47 years, 169 days (17,336 days);Jan 20, 2017;55 years, 169 days (20,258 days)
9 ;alive;55 years, 158 days (20,247 days)
10 45;Donald Trump;Jun 14, 1946;Jan 20, 2017;70 years, 220 days (25,788 days);;;alive;70 years, 209 days (25,777 days)

```

3. Wijzig in de vier datumvelden het formaat van `Mon d, yyyy` naar `d Mon yyyy`. We krijgen dan

```

1 1;George Washington;22 Feb 1732;30 Apr 1789;57 years, 67 days (20,887 days);4 Mar 1797;65 years, 10 days (23,752 days)
2 ;2 years, 285 days (1,015 days);14 Dec 1799;67 years, 295 days (24,767 days)
3 2;John Adams;30 Oct 1735;4 Mar 1797;61 years, 125 days (22,406 days);4 Mar 1801;65 years, 125 days (23,866 days);25
4 years, 122 days (9,253 days);4 Jul 1826;90 years, 247 days (33,119 days)
5 3;Thomas Jefferson;13 Apr 1743;4 Mar 1801;57 years, 325 days (21,144 days);4 Mar 1809;65 years, 325 days (24,066 days)
6 ;17 years, 122 days (6,331 days);4 Jul 1826;83 years, 82 days (30,397 days)
7 4;James Madison;16 Mar 1751;4 Mar 1809;57 years, 353 days (21,172 days);4 Mar 1817;65 years, 353 days (24,094 days);19
8 years, 116 days (7,056 days);28 Jun 1836;85 years, 104 days (31,150 days)
9 5;James Monroe;28 Apr 1758;4 Mar 1817;58 years, 310 days (21,494 days);4 Mar 1825;66 years, 310 days (24,416 days);6
10 years, 122 days (2,313 days);4 Jul 1831;73 years, 67 days (26,729 days)
11 ...

```

4. Zorg ervoor dat in de vier velden die een tijdsperiode aanduiden enkel de duur van de tijdsperiode uitgedrukt in dagen behouden blijft (de tijdsaanduiding die initieel tussen ronde haakjes stond). Hierbij moeten ook de suffix `_days` en de komma's tussen de cijfers van de tijdsduur uitgedrukt in dagen verwijderd worden. We krijgen dan

```

1 | 1;George Washington;22 Feb 1732;30 Apr 1789;20887;4 Mar 1797;23752;1015;14 Dec 1799;24767
2 | 2;John Adams;30 Oct 1735;4 Mar 1797;22406;4 Mar 1801;23866;9253;4 Jul 1826;33119
3 | 3;Thomas Jefferson;13 Apr 1743;4 Mar 1801;21144;4 Mar 1809;24066;6331;4 Jul 1826;30397
4 | 4;James Madison;16 Mar 1751;4 Mar 1809;21172;4 Mar 1817;24094;7056;28 Jun 1836;31150
5 | 5;James Monroe;28 Apr 1758;4 Mar 1817;21494;4 Mar 1825;24416;2313;4 Jul 1831;26729
6 | ...

```

5. Sorteert de lijst van presidenten volgens dalende levensduur (waardoor de langst levende presidenten helemaal bovenaan komen te staan). Zorg er evenwel voor dat de volgnummers in oplopende volgorde blijven staan, te beginnen vanaf 1. We krijgen dan

```

1 | 1;Gerald Ford;14 Jul 1913;9 Aug 1974;22306;20 Jan 1977;23201;10932;26 Dec 2006;34133
2 | 2;Ronald Reagan;6 Feb 1911;20 Jan 1981;25551;20 Jan 1989;28473;5615;5 Jun 2004;34088
3 | 3;George H. W. Bush;12 Jun 1924;20 Jan 1989;23598;20 Jan 1993;25059;8757;alive;33814
4 | 4;Jimmy Carter;1 Oct 1924;20 Jan 1977;19104;20 Jan 1981;20565;13140;alive;33703
5 | 5;John Adams;30 Oct 1735;4 Mar 1797;22406;4 Mar 1801;23866;9253;4 Jul 1826;33119
6 | ...

```

Opgave 2 (uencode)

(10 pt)

Het UUCP systeem voor elektronische mail — een van de oudste emailprotocollen — liet enkel toe om tekstuele informatie uit te wisselen tussen computersystemen. Om toch binaire bestanden te kunnen versturen, zocht men een manier om binaire data om te zetten naar tekst. Hiervoor gebruikte men een techniek die **uencoding** genoemd wordt (afgeleid van *Unix-to-Unix encoding*). De codering van een binair bestand zag er dan bijvoorbeeld als volgt uit:

```

begin 644 bestand.txt
M22!F965L('9E<GD@<W1R;VYG;'D@86)0=70@>6]U(&10:6YG(&1U='DN(%=0
M=6QD('EO=2!G:79E(&UE(&$@;&ET=&QE(&U0<F4@9&]C=6UE;G1A=&EO;B!A
M8F]U="!Y;W5R(')E861I;F<@:6X@1G)E;F-H/R!)(%M(&=L860@>6]U(&%R
M92!H87!P>2#B@)0@8G5T($D@;F5V97(@8F5L:65V92!M=6-H(&EN(&AA<'!I
M;F5S<RX@22!N979E<B!B96QI979E(&EN(&UI<V5R>2!E:71H97(N(%1H;W-E
M(&%R92!T:&EN9W,@>6]U('E92!O;B!T:&4@<W1A9V4@;W(@=&AE('C<F5E
M;B!O<B!T:&4@<'I;G1E9"!P86=E<RP@=&AE>2!N979E<B!R96%L;'D@:&%P
3<&5N('10('EO=2!I;B!L:69E+C P
`
end

```

Een *uencoded* bestand begint met één regel koptekst, en eindigt met twee regels voettekst. Alle tussenliggende regels bevatten enkel karakters met een ASCII-waarde tussen 32 en 95, waardoor alle symbolen die in deze codering gebruikt worden afdrukbaar zijn.

Opgave

Elke regel van het tekstbestand `uencoded.txt` bevat een reeks karakters die enkel bestaan uit hoofdletters, cijfers en leestekens (met ASCII-waarde tussen 33 en 95). Er komen dus geen kleine letters en spaties voor in de reeks karakters. Het aantal karakters in een reeks kan variëren van regel tot regel. Na de reeks karakters volgt nog een spatie en een woord dat enkel bestaat uit kleine letters. Gevraagd wordt:

- Bepaal reguliere expressies voor elk van onderstaande verzamelingen. Daarbij staat \mathcal{U} voor de verzameling van alle mogelijke reeksen karakters zoals die voorkomen aan het begin van de regels uit het bestand `uencoded.txt`.

(a) $\alpha = \{u \in \mathcal{U} \mid u \text{ bevat een punt, een vraagteken en een uitroepteken} \}$

voorbeelden: `I?D.(W1PT0']E~A&*3"B=/>?1~H19QK'!^3&BDY8R[3-/OZ\CEIP/ motorcycle` $\in \alpha$,
`~>]/G7F\D31KN^DDFMF1S&]W-\M&'F5]ZPF&{YE4*~>97]_2-PV/)6K5 lunula` $\notin \alpha$

(b) $\beta = \{u \in \mathcal{U} \mid u \text{ bevat hoogstens twee opeenvolgende leestekens}\}$

voorbeelden: *OIC\$M\$R3D6JONVOIV=?N{-5N&'9I:V.?L}/3.@ICT:1TCG3^2=)F(>6 hydronium $\in \beta$,
5AL(*K>*2LZ>?34-\I~NV}R7PC"39<|3'#Z*Y}*3.>9.UJJC{ET\$.FI~M3 adherend $\notin \beta$

(c) $\gamma = \{u \in \mathcal{U} \mid u \text{ begint met een hoofdletter die exact drie keer voorkomt in } u\}$

voorbeelden: X^ [<*'??XBXK<:B7#; }VO|<|K!(P|HW}{10@\$JK_})*.5H"Y&^A)D\$QS snaking $\in \gamma$,
R1>7S_\R6[_KRLY]ZM}OD,7OH?.R(F00H^PEB[&X'R(!B_0<N|(*'&**0TB; penni $\notin \gamma$

(d) $\delta = \{u \in \mathcal{U} \mid \text{in } u \text{ worden ronde haakjes alternerend geopend en gesloten, en elk geopend rond haakje moet verderop ook terug afgesloten worden}\}$

voorbeelden: U1~;15|ES!' "H(HUQJE)_8T,R[(3 F-F,;/XMJ[9W)QPF36R*6F6 quassins $\in \delta$,
D8FZ!H@C:QQV57Z)4HLP-^R5V0.' :3Q>(#N2PG@*FDO)[I1_;0Y]V={QPD= bitt $\notin \delta$

Gebruik een commando uit de **grep** familie om enkel die regels van het bestand **uuencoded.txt** te selecteren die behoren tot de opgegeven verzameling. Vermeld in je antwoordbestand voor elke verzameling het gebruikte selectiecommando, en geef telkens ook aan hoeveel regels je gevonden hebt.

2. Beschouw de verzamelingen α , β , γ en δ zoals hierboven gedefinieerd. Gebruik nu deze verzamelingen om op de volgende manier een boodschap bestaande uit vier woorden te achterhalen:

(a) het eerste woord staat op de unieke regel uit de verzameling $\alpha \cap \beta$

(b) het tweede woord staat op de unieke regel uit de verzameling $\beta \cap \gamma$

(c) het derde woord staat op de unieke regel uit de verzameling $\gamma \cap \delta$

(d) het vierde woord staat op de unieke regel uit de verzameling $\delta \cap \alpha$

Vermeld in je antwoordbestand de gevonden woorden, samen met het Unix commando (of de commandosequentie) dat je gebruikt hebt om elk van deze woorden te vinden.

Opgave 3 (uitslover)

(10 pt)

Er zijn slechts twaalf mensen die de vier grote jaarlijkse Amerikaanse entertainmentprijzen gewonnen hebben: Emmy, Grammy, Oscar en Tony. Het winnen van deze vier prijzen wordt afgekort als EGOT (of soms als GATE als men de officiële naam Academy Award gebruikt voor de Oscar) en men spreekt ook van de Grand Slam van de showbusiness. Deze prijzen eren uitstekende prestaties op gebied van televisie, muziek (of andere audio-opnames), film en theater.

NAAM	EGOT VOLTOOID	EMMY	GRAMMY	OSCAR	TONY
Richard Rodgers	1962 (17 jaar)	1962	1960	1945	1950
Helen Hayes	1977 (45 jaar)	1953	1977	1932	1947
Rita Moreno	1977 (16 jaar)	1977	1972	1961	1975
John Gielgud	1991 (30 jaar)	1991	1979	1981	1961
Audrey Hepburn	1994 (41 jaar)	1993	1994	1953	1954
Marvin Hamlisch	1995 (23 jaar)	1995	1974	1973	1976
Jonathan Tunick	1997 (20 jaar)	1982	1988	1977	1997
Mel Brooks	2001 (34 jaar)	1967	1998	1968	2001
Mike Nichols	2001 (40 jaar)	2001	1961	1967	1964
Whoopi Goldberg	2002 (17 jaar)	2002	1985	1990	2002
Scott Rudin	2012 (28 jaar)	1984	2012	2007	1994
Robert Lopez	2014 (10 jaar)	2008	2012	2014	2004

Het acroniem EGOT werd voor het eerst gebruikt door acteur Philip Michael Thomas toen hij onthulde dat hij de ambitie had om deze vier prijzen te winnen. Tot op heden werd hij echter nog voor geen enkele van de prijzen genomineerd, maar hij won wel een People's Choice Award en een nominatie voor de Golden Globe. Drie andere personen — Barbra Streisand, Liza Minelli en James Earl Jones — hebben ook de vier prijzen gewonnen, maar één daarvan werd buiten competitie uitgereikt als een soort ere-onderscheiding (Streisand's Tony, Minelli's Grammy en Jones' Oscar). Van de twaalf winnaars zijn er ook nog twee — Marvin Hamlisch and Richard Rodgers — die de Pulitzerprijs in de wacht gesleept hebben.

De volgende personen hebben elk drie van de vier grote entertainmentprijzen gewonnen in competitieve categorieën. Het winnen van drie van de vier prijzen wordt ook wel eens 3GOT genoemd (kwinkslag naar de term EGOT), waarbij de vier combinaties afzonderlijk benoemd worden als EGO, TOE, GOT en GET.

GEEN TONY (EGO)	GEEN GRAMMY (TOE)	GEEN EMMY (GOT)	GEEN OSCAR (GET)
John Addison	Jack Albertson	Henry Fonda	Harry Belafonte
Julie Andrews	Anne Bancroft	Oscar Hammerstein II	Leonard Bernstein
Burt Bacharach	Ingrid Bergman	Elton John	Martin Charnin
Alan Bergman	Shirley Booth	Alan Jay Lerner	Cy Coleman
Marilyn Bergman	Ralph Burns	Andrew Lloyd Webber	Fred Ebb
George Burns	Ellen Burstyn	Frank Loesser	Anne Garefino
Cher	Melvyn Douglas	Alan Menken	Julie Harris
Ray Dolby	Bob Fosse	Tim Rice	James Earl Jones
Michael Giacchino	Jeremy Irons	Stephen Sondheim	John Kander
James Moll	Liza Minnelli	Jule Styne	Cyndi Lauper
Randy Newman	Thomas Mitchell		John McDaniel
Sid Ramin	Al Pacino		Cynthia Nixon
Martin Scorsese	Christopher Plummer		Trey Parker
Barbra Streisand	Vanessa Redgrave		Marc Shaiman
Peter Ustinov	Jason Robards		Matt Stone
John Williams	Geoffrey Rush		Charles Strouse
Robin Williams	Paul Scofield		Lily Tomlin
Kate Winslet	Maggie Smith		Dick Van Dyke
	Maureen Stapleton		James Whitmore
	Peter Stone		
	Jessica Tandy		
	Tony Walton		

Opgave

Gegeven zijn vier tekstbestanden `emmy.txt`, `grammy.txt`, `oscar.txt` en `tony.txt`, die elk een lijst van personen bevatten die de gelijknamige prijs gewonnen hebben. Elke regel van die bestanden bevat de naam van een persoon, gevolgd door een tab en het jaartal waarin die persoon de prijs gewonnen heeft. Je mag ervan uitgaan dat een persoon maar één keer in het bestand voorkomt (indien een persoon meerdere keren een bepaalde prijs gewonnen heeft, dan staat in het bestand bijvoorbeeld enkel het jaartal waarin de prijs voor het eerst gewonnen werd).

Gevraagd wordt om een `bash` shell script `uitslover` te schrijven dat op basis van de inhoud van deze vier tekstbestanden een overzicht moet uitschrijven naar standaard uitvoer met alle personen die twee, drie of vier van de prijzen gewonnen hebben. Het shell script moet er hierbij van uitgaan dat de vier bestanden in dezelfde directory als het shell script staan. Het overzicht is een tabel met in de eerste kolom de naam van de persoon en in de volgende kolommen de jaartallen waarin de prijzen

gewonnen werden. De personen moeten in alfabetische volgorde in de tabel opgelijst worden. Alle kolommen moeten van elkaar gescheiden worden door één enkele spatie. Aan de tabel moet een hoofding toegevoegd worden, waarbij de kolommen vaste namen hebben die je kan afleiden uit onderstaande voorbeelden. Elke kolomhoofding moet onderlijnd worden met koppeltekens zoals aangegeven in de voorbeelden.

Aan het shell script mogen geen argumenten doorgegeven worden, maar het shell script moet wel de volgende vier opties ondersteunen:

- optie `-E`: neem winnaars van een Emmy Award op in het overzicht
- optie `-G`: neem winnaars van een Grammy Award op in het overzicht
- optie `-O`: neem winnaars van een Oscar op in het overzicht
- optie `-T`: neem winnaars van een Tony Award op in het overzicht

Enkel de personen die alle gevraagde prijzen gewonnen hebben mogen in het overzicht opgenomen worden, en het overzicht mag enkel kolommen bevatten voor de gevraagde prijzen. Merk op dat de kolommen met de jaartallen waarin de prijzen gewonnen werden steeds in een vaste volgorde moeten opgelijst worden (Emmy–Grammy–Oscar–Tony), ongeacht de volgorde waarin de opties aan het shell script doorgegeven worden. Bovendien moet het shell script de volgende foutafhandeling voorzien:

- indien het shell script niet de gepaste opties meekrijgt (enkel ondersteuning voor de opties `-E`, `-G`, `-O` en `-T`), een optieletter meerdere keren wordt meegegeven, of er worden argumenten meegegeven aan het shell script dan moet de gepaste boodschap (zie onderstaande voorbeeldsessie) naar standaard error uitgeschreven worden, en moet het shell script eindigen met exit status 1
- indien aan het shell script niet minstens twee opties meegegeven worden, dan moet de gepaste boodschap (zie onderstaande voorbeeldsessie) naar standaard error uitgeschreven worden, en moet het shell script eindigen met exit status 2

Tip: wellicht kan je voor het schrijven van het shell script handig gebruik maken van het Unix commando `join`.

Voorbeeld

Onderstaande voorbeeldsessie geeft aan hoe het shell script `uitslover` moet kunnen gebruikt worden. Hierbij gaan we ervan uit dat de tekstbestanden `emmy.txt`, `grammy.txt`, `oscar.txt` en `tony.txt` zich in de huidige directory bevinden.

```
$ uitslover -EGOT
NAAM ACTEUR      EMMY      GRAMMY    OSCAR     TONY
-----
Brooks Mel       1967      1998      1968      2001
Gielgud John     1991      1979      1981      1961
Goldberg Whoopi 2002      1985      1990      2002
Hamlisch Marvin   1995      1974      1973      1976
Hayes Helen     1953      1977      1932      1947
Hepburn Audrey  1993      1994      1953      1954
Lopez Robert    2008      2012      2014      2004
Moreno Rita     1977      1972      1961      1975
Nichols Mike    2001      1961      1967      1964
Rodgers Richard 1962      1960      1945      1950
Rudin Scott     1984      2012      2007      1994
Tunick Jonathan 1982      1988      1977      1997
$ echo $?
0
```

```

$ uitslover -TOE
NAAM ACTEUR      EMMY      OSCAR      TONY
-----
Brooks Mel       1967      1968      2001
Gielgud John    1991      1981      1961
Goldberg Whoopi 2002      1990      2002
Hamlisch Marvin  1995      1973      1976
Hayes Helen     1953      1932      1947
Hepburn Audrey  1993      1953      1954
Lopez Robert    2008      2014      2004
Moreno Rita     1977      1961      1975
Nichols Mike    2001      1967      1964
Rodgers Richard 1962      1945      1950
Rudin Scott     1984      2007      1994
Tunick Jonathan 1982      1977      1997
$ uitslover -AGOT
Syntaxis: uitslover -EGOT
$ echo $?
1
$ uitslover -EGOT argument
Syntaxis: uitslover -EGOT
$ echo $?
1
$ uitslover -T
uitslover: minder dan twee awards opgegeven
$ echo $?
2

```

Opgave 4 (rijmschema)

(10 pt)

In deze opgave willen we het rijmschema van een gegeven gedicht bepalen. Hiervoor beschikken we over een tekstbestand waarin woorden aan hun **uitspraak** gekoppeld worden. Het bestand begint met een aantal commentaarregels die starten met een hekje (#). Deze mogen genegeerd worden. Daarna staat op elke regel een woord (in hoofdletters) en zijn uitspraak. De regel

```
PROGRAMMING P R OW1 G R AE2 M IHO NG
```

geeft bijvoorbeeld aan hoe het woord **programming** moet uitgesproken worden. Het woord wordt hierbij gescheiden van zijn uitspraak door één of meer witruimtekarakters (het woord zelf bevat geen witruimtekarakters), en de uitspraak wordt beschreven door een reeks fonemen die telkens van elkaar gescheiden worden door één spatie. In totaal wordt er voor het beschrijven van de uitspraak gebruik gemaakt van 39 fonemen (15 klinkers en 24 medeklinkers; fonemen bestaan enkel uit hoofdletters). Elke **klinker** heeft drie mogelijke klemtonen die worden aangegeven door een cijfer op het einde van het foneem (0 indien geen nadruk, 1 indien primaire nadruk en 2 indien secundaire nadruk). In totaal worden er dus 69 symbolen gebruikt om de uitspraak weer te geven.

```

AA odd      AA D
AE at       AE T
AH hut      HH AH T
AO ought    AO T
AW cow      K AW
AY hide     HH AY D
B be        B IY
CH cheese   CH IY Z
D dee       D IY
DH thee     DH IY
EH Ed       EH D
ER hurt     HH ER T
EY ate      EY T
F fee       F IY
G green     G R IY N

```

HH	he	HH	IY
IH	it	IH	T
IY	eat	IY	T
JH	gee	JH	IY
K	key	K	IY
L	lee	L	IY
M	me	M	IY
N	knee	N	IY
NG	ping	P	IH NG
OW	oat	OW	T
OY	toy	T	OY
P	pee	P	IY
R	read	R	IY D
S	sea	S	IY
SH	she	SH	IY
T	tea	T	IY
TH	theta	TH	EY T AH
UH	hood	HH	UH D
UW	two	T	UW
V	vee	V	IY
W	we	W	IY
Y	yield	Y	IY L D
Z	zee	Z	IY
ZH	seizure	S	IY ZH ER

Er zijn echter ook woorden die meerdere **alternatieve uitspraken** hebben. In het tekstbestand wordt elk van deze alternatieven weergegeven op een afzonderlijke regel. De alternatieven worden genummerd door na het woord in hoofdletters een getal tussen ronde haakjes te plaatsen. Zo geven de regels

DONKEY	D	AA1	NG	K	IYO
DONKEY(2)	D	AO1	NG	K	IYO

twee alternatieve uitspraken aan van het woord **donkey**.

Opgave

De **uitgang** van een uitspraak wordt bepaald als de suffix (substring die loopt tot het einde van de uitspraak) die begint bij de laatste klinker met primaire of secundaire nadruk, of aan het begin van de uitspraak als er geen benadrukte klinkers zijn. Het resterende deel (de prefix voor die laatste klinker) wordt de **stam** van de uitspraak genoemd. De uitspraak P R OW1 G R AE2 M IHO NG heeft dus stam P R OW1 G R en uitgang AE2 M IHO NG.

We zeggen dat twee woorden **rijmen** als hun uitspraak dezelfde uitgang heeft. Hierbij houden we voor deze opgave enkel rekening met de eerste alternatieve uitspraak van een woord. Alle alternatieven met een getal tussen ronde haakjes na het woord worden met andere woorden genegeerd. Bovendien houden we bij het vergelijken van de uitgangen geen rekening met mogelijke klemtoonverschillen in de uitgang. Voor het woord **programming** gebruiken we bij het bepalen van de rijmwoorden dus de uitgang AE M IH NG (zonder klemtonen 0, 1 en 2).

Gevraagd wordt om een **bash** shell script **rijmschema** te schrijven waaraan twee argumenten moeten doorgegeven worden: *i*) een tekstbestand met enkele regels van een gedicht en *ii*) een tekstbestand waarin woorden aan hun uitspraak gekoppeld worden. Het shell script moet naar standaard uitvoer een regel uitschrijven met daarin het rijmschema dat gebruikt wordt in het gedicht. Dit rijmschema bestaat uit een reeks hoofdletters die aangeven wat de uitgang is van het laatste woord op de opeenvolgende regels van het gedicht. De uitgang van het laatste woord op de eerste regel wordt altijd aangeduid met de hoofdletter A. Als het laatste woord op de volgende regel een uitgang heeft die we nog niet zijn tegengekomen, dan wordt die aangeduid met de eerstvolgende ongebruikte hoofdletter van het alfabet. Anders wordt die aangeduid met de hoofdletter die we al hadden toegekend aan die uitgang. Om het laatste woord van een regel te bepalen, gebruiken we de definitie dat de woorden van een regel gevormd worden door de langst mogelijke opeenvolging van letters. Regels van het gedicht die geen woorden bevatten moeten genegeerd worden bij het bepalen van het rijmschema.

Voor deze opgave is het niet nodig dat het shell script foutafhandeling voorziet voor de argumenten die aan het shell script worden doorgegeven. Je mag er dus altijd van uitgaan dat er twee geldige bestandslocaties aan het shell script doorgegeven worden. Bovendien mag het shell script ervan uitgaan dat het bestand met de uitspraken de uitspraak vastlegt van alle laatste woorden op alle regels van het bestand dat de tekst van het gedicht bevat.

Voorbeeld

Onderstaande voorbeeldsessie geeft aan hoe het shell script `rijmschema` moet kunnen gebruikt worden. Hierbij gaan we ervan uit dat de tekstbestanden `gedicht01.txt`, ...`gedicht05.txt`, en het tekstbestand `c06d` zich in de huidige directory bevinden.

```
$ cat gedicht01.txt
A bather whose clothing was screwed
By winds that left her quite nude
Saw a man come along
And unless we are wrong
You expected this line to be lewd.
$ rijmschema gedicht01.txt c06d
AABBA
$ echo $?
0
$ rijmschema gedicht02.txt c06d
AABBA
$ rijmschema gedicht03.txt c06d
ABCBBB
$ rijmschema gedicht04.txt c06d
AABCCC
$ rijmschema gedicht05.txt c06d
ABAAB
```