

InfoTrend computer coding of open-ended responses

By

David P. Fan

dfan@umn.edu

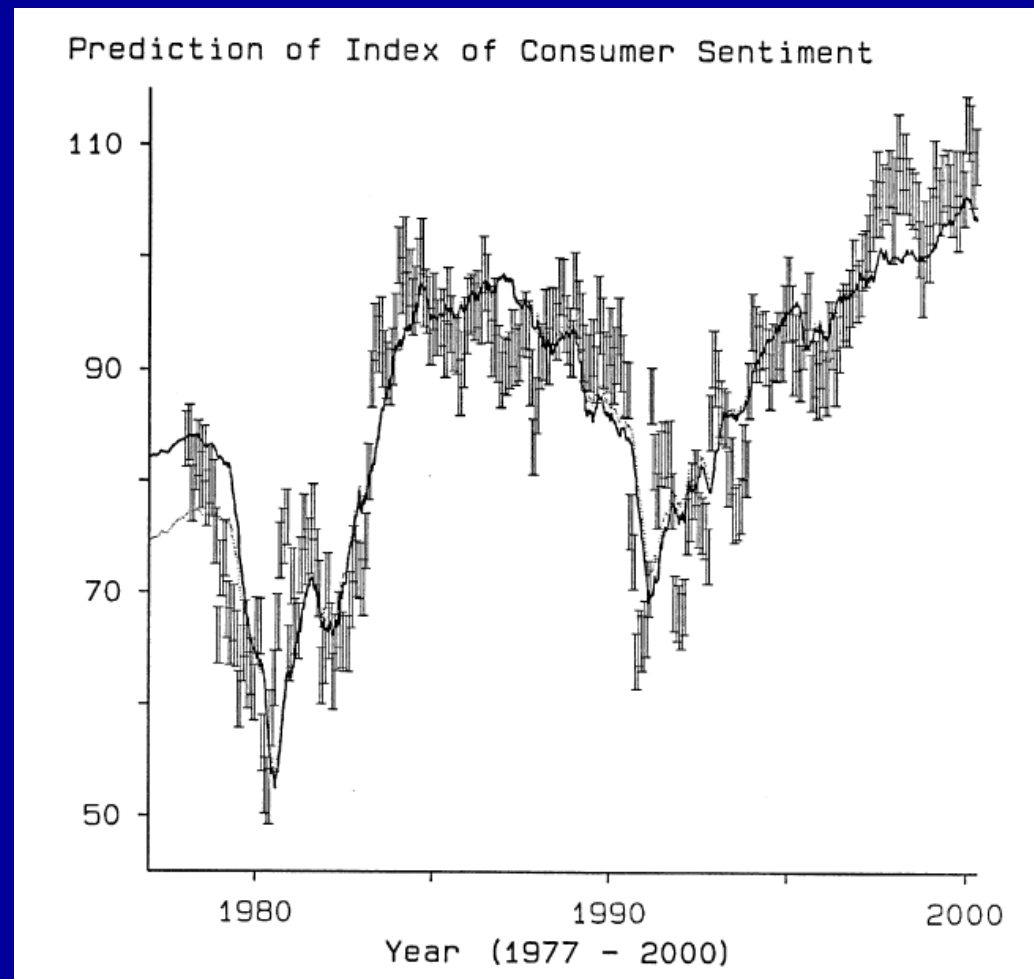
Background

Biology	Opinion/Behavior
Hormones (e.g. insulin) are chemical messages in identical copies	News media consist of messages in identical copies
Sent by one group of cells (e.g. islet cells)	Sent by one group of people called publishers
To influence other cells in the body	To influence other people in the population
The parallel led to transferal of biological logic to the ideodynamic model for opinion formation	

Ideodynamic prediction of the Index of Consumer Sentiment

- Predictor: news coverage
 - 45,145 Washington Post and Associated Press stories on the economy
 - Scored using the InfoTrend computer system as good or bad for the economy
- Use media alone in the differential equation, non-autoregressive model of ideodynamics to predict monthly values for the Index of Consumer Sentiment constructed from 5 survey questions

Goal: Predict Line to Match Survey Data
($R^2=0.83$; Symbols= 95% confidence for
Sentiment; Fan and Cook, J. Math. Sociol.(2003)



Overview of InfoTrend system

- A human analyst uses theory to specify ideas to be identified in text
- The analyst writes computer algorithms in the InfoTrend format to specify how the ideas are to be identified in the text
 - The InfoTrend format mimics human reasoning, and
 - Is thus easy for humans to read, understand and revise
 - Ideas can be identified in multiple steps
 - E.g. for economic text
 - The text was first filtered to be about the U.S. economy
 - The filtered text was then coded as being good and bad for the economy
- It is easy to obtain the same degree of agreement in machine-human and human-human comparisons

Outline of InfoTrend system

- The software uses pre-assigned flags to divide input text into documents with paragraphs that can be assigned
 - To be sentences, or
 - To begin with pre-assigned flags: Flag for new paragraphs for ANES 1992 open ends was prompts e.g. (AE) = “anything else”
- The user writes algorithms to identify paragraphs containing the specified ideas
- The user examines the results of algorithm action on screen and can iteratively modify the theory behind the ideas and/or the algorithms specifying the ideas
- The software applies the algorithms to the entire corpus of text after the user deems the algorithms to be satisfactory
 - The output is the count of user specified ideas in each paragraph
 - A paragraph can be assigned to have more than one idea
 - An idea can be assigned to a document as the count of the paragraphs in the document with the idea

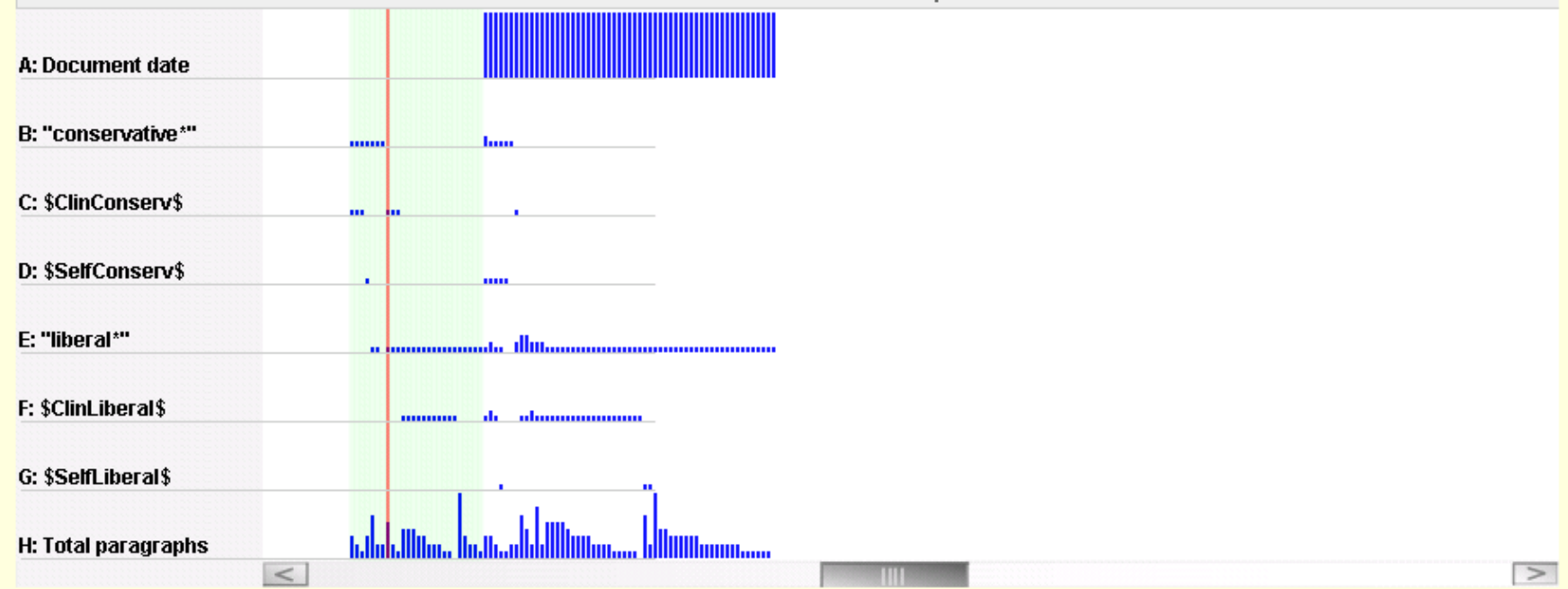
Example of InfoTrend analysis

- ANES 1992 Like/Dislike Clinton text
- The user decided on a group of related ideas to be assigned. The ideas were:
 - Clinton is liberal
 - Clinton is conservative
 - Respondent is liberal
 - Respondent is conservative
- The user examined text enriched for the idea group to look for text containing the ideas
 - For the above ideas, the user selected responses with the words: “conservative” or “liberal”

Looking at text

- (Not use standard display of documents)
 - (List of document fragments with only a few tens visible at a time)
- Use InfAlign map to display 100s of documents together on screen (next slide)
 - Each document is shown as a stack of bar graphs displaying user selected traits
 - Remove document labels, make bars narrow, squeeze together
 - E.g. 8 bars for conservative/liberal open ends in following slide
 - Document Date bar:
 - All Like documents assigned to 1/3/1992 (height is floor)
 - All Dislike documents assigned to 1/4/1992 (height is ceiling)
 - Background color changes on date change
 - “Conservative” bar counts paragraphs with the word “conservative”
 - \$ClinConserv\$ bar counts “Clinton is conservative” paragraphs
 - \$SelfConserv\$ bar counts “Respondent is conservative” paragraphs
 - Equivalent set of 3 bars for liberal
 - Total Paragraphs bar counts the paragraphs in the document
 - Hover computer cursor over stack of bars for a document
 - Information about that document is shown at top of map
 - Clicking on a stack opens window with full document

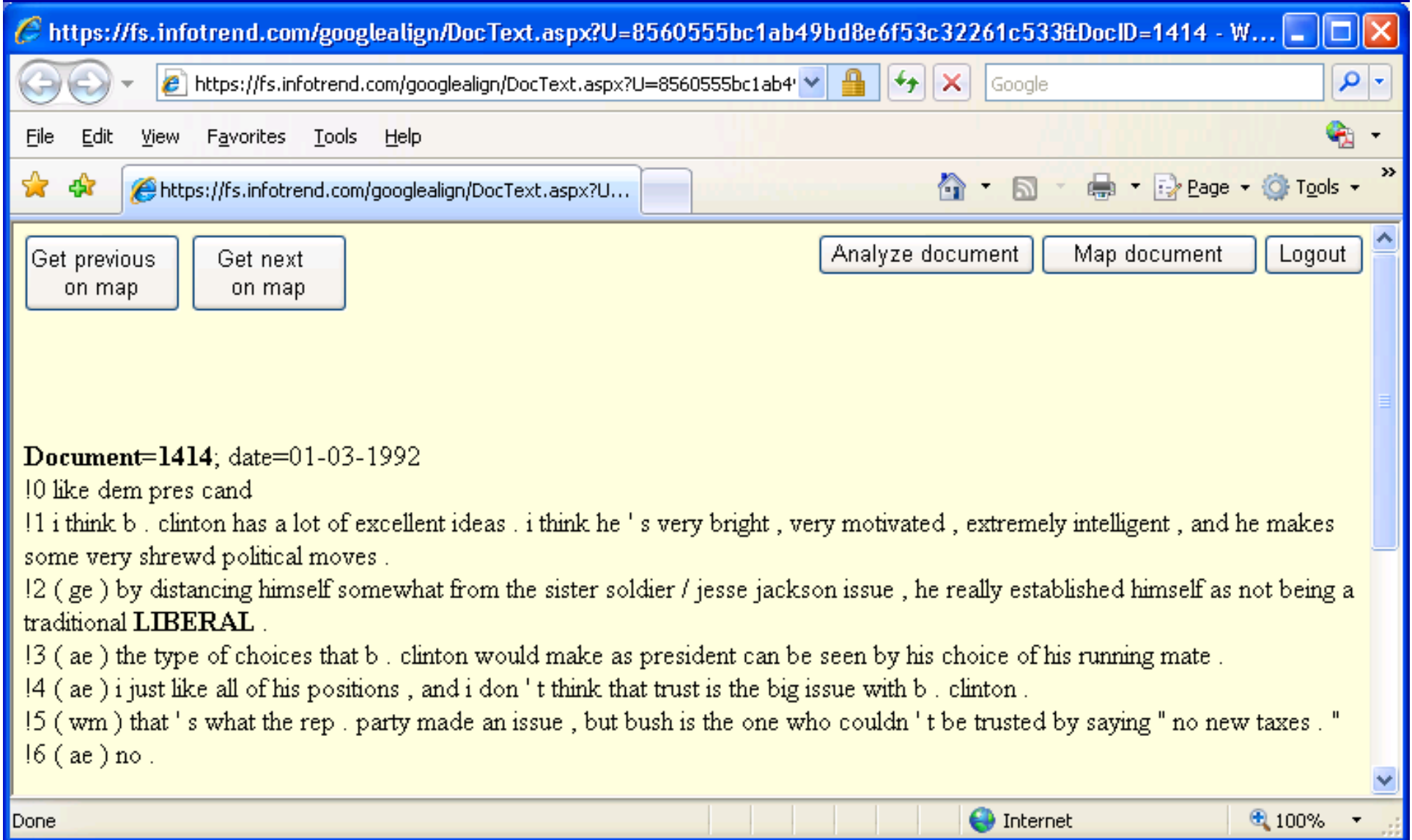
ID=1414 01-03-1992 C:1 E:1 H:7 like dem pres cand



Mapping Statistics

<u>Mapping Conditions</u>	<u>Matches to Conditions</u>	<u>Data Range</u>	<u>Bar Height Range</u>
Date range	01/03/1992-01/04/1992		
Documents in date range	2665		
Documents after filtering	83		
Filtered documents mapped	83		
B: "conservative*"	13	0-2	0-12
C: \$ClinConserv\$	7	0-1	0-12
D: \$SelfConserv\$	6	0-1	0-12

Window opened by clicking on the red bar on the InfAlign map of the previous slide



The screenshot shows a web browser window with the following elements:

- Address Bar:** <https://fs.infotrend.com/googlealign/DocText.aspx?U=8560555bc1ab49bd8e6f53c32261c533&DocID=1414>
- Navigation:** Back, Forward, Home, Stop, Refresh, and a search box with "Google" text.
- Menu Bar:** File, Edit, View, Favorites, Tools, Help.
- Page Navigation:** "Get previous on map" and "Get next on map" buttons on the left; "Analyze document", "Map document", and "Logout" buttons on the right.
- Document Content:**

Document=1414; date=01-03-1992
10 like dem pres cand
11 i think b . clinton has a lot of excellent ideas . i think he ' s very bright , very motivated , extremely intelligent , and he makes some very shrewd political moves .
12 (ge) by distancing himself somewhat from the sister soldier / jesse jackson issue , he really established himself as not being a traditional **LIBERAL** .
13 (ae) the type of choices that b . clinton would make as president can be seen by his choice of his running mate .
14 (ae) i just like all of his positions , and i don ' t think that trust is the big issue with b . clinton .
15 (wm) that ' s what the rep . party made an issue , but bush is the one who couldn ' t be trusted by saying " no new taxes . "
16 (ae) no .
- Status Bar:** Done, Internet, 100% zoom.

- Idea section of InfoTrend algorithm
 - User declares a set of ideas on the left margin with ideas indicated
 - By words (any strings of characters) in { } symbols underneath, or
 - By combining other ideas using rules shown below
 - Ideas for which codes are generated are labeled cSCORED along with a description of the code e.g. Clinton is conservative

```

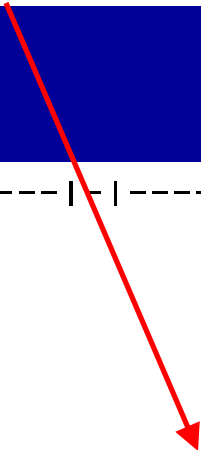
----->-----|----->-----|----->-----
Barrier
{ . }
Negation
{ hurt }
FwdNeg
{ less } { never } { no } {not }
CancelFwdNeg
{ more }
Self
{ i }
Clinton
{ clinton } { he }
NotSelf
{agree}
Conservative
{ conservative}
Liberal
{liberal}
ClinConserv      1 cSCORED      'Clinton is conservative'
ClinLiberal      2 cSCORED      'Clinton is liberal'
SelfConserv      3 cSCORED      'I am conservative'
SelfLiberal      4 cSCORED      'I am liberal'
----->-----|----->-----|----->-----

```

- Rule section of InfoTrend algorithm
 - User writes rules for how ideas combine pair wise to give other ideas based on the text equivalent
 - E.g. rules can combine two ideas into one or split one idea into two when an idea has more than one useful meaning
 - Rules can work in sequence to give progressively more complex ideas
 - Rules are numbered in the column at the far right
 - (Description and use of rules is seen in following slides)

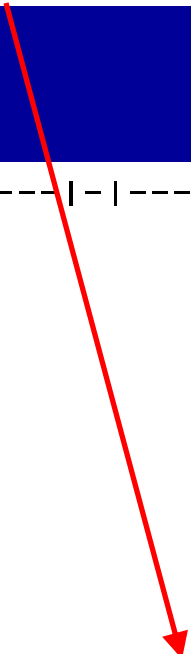
Negation	E	20	Conservative	Liberal	1
Negation	E	20	Liberal	Conservative	2
CancelFwdNeg	B	10	FwdNeg	DELETED	3
			CancelFwdNeg	DELETED	4
FwdNeg	A	20	Conservative	Liberal	5
FwdNeg	A	20	Liberal	Conservative	6
NotSelf	B	10	Self	DELETED	7
Self	A	30	Liberal	SelfLiberal	8
Self	A	30	Conservative	SelfConserv	9
Clinton	A	30	Liberal	ClinLiberal	10
Clinton	A	30	Conservative	ClinConserv	11
Conservative	E	-1	ClinLiberal	DELETED	R 12

Rule 6 below	Expanded form for human reading
FwdNeg	If an idea of the FwdNeg group e.g. {not }
A	Has to its right as indicted by the letter A
20	Within 20 characters of text
Liberal	An idea of the Liberal class e.g. {liberal}
Conservative	Then change the ideas to the idea Conservative
6	(6 is the rule number)



Negation	E	20	Conservative	Liberal	1
Negation	E	20	Liberal	Conservative	2
CancelFwdNeg	B	10	FwdNeg	DELETED	3
			CancelFwdNeg	DELETED	4
FwdNeg	A	20	Conservative	Liberal	5
FwdNeg	A	20	Liberal	Conservative	6
NotSelf	B	10	Self	DELETED	7
Self	A	30	Liberal	SelfLiberal	8
Self	A	30	Conservative	SelfConserv	9
Clinton	A	30	Liberal	ClinLiberal	10
Clinton	A	30	Conservative	ClinConserv	11
Conservative	E	-1	ClinLiberal	DELETED	R 12

Rule 11 below	Expanded form for human reading
Clinton	If an idea of the Clinton group e.g. { he }
A	Has to its right as indicted by the letter A
30	Within 30 characters of text
Conservative	An idea of the Conservative class e.g. from Rule 6
ClinConserv	Then change the ideas to 'Clinton is conservative'
11	(11 is the rule number)



```

-----|-----|-----|-----|-----|
Negation      E      20      Conservative      Liberal      1
Negation      E      20      Liberal           Conservative  2
CancelFwdNeg  B      10      FwdNeg           DELETED      3
              B      10      CancelFwdNeg     DELETED      4
FwdNeg        A      20      Conservative     Liberal      5
FwdNeg        A      20      Liberal          Conservative  6
NotSelf       B      10      Self            DELETED      7
Self          A      30      Liberal         SelfLiberal   8
Self          A      30      Conservative    SelfConserv   9
Clinton       A      30      Liberal         ClinLiberal  10
Clinton       A      30      Conservative    ClinConserv  11
Conservative  E      -1      ClinLiberal     DELETED      R      12
-----|-----|-----|-----|-----|

```

Application of InfoTrend rules

- Visualization of rules in action: Machine writes to screen (next slide)
 - The text of a paragraph
 - Beneath that, a text equivalent for the paragraph giving ideas corresponding to words and distances between them in characters of text
 - Beneath that, a map of rule usage
 - Rules 6 and 11 as described above are used in sequence
- The user changes the ideas, words, and rules to improve assignment precision
- When the user is satisfied, the machine applies the instructions to the entire set of documents

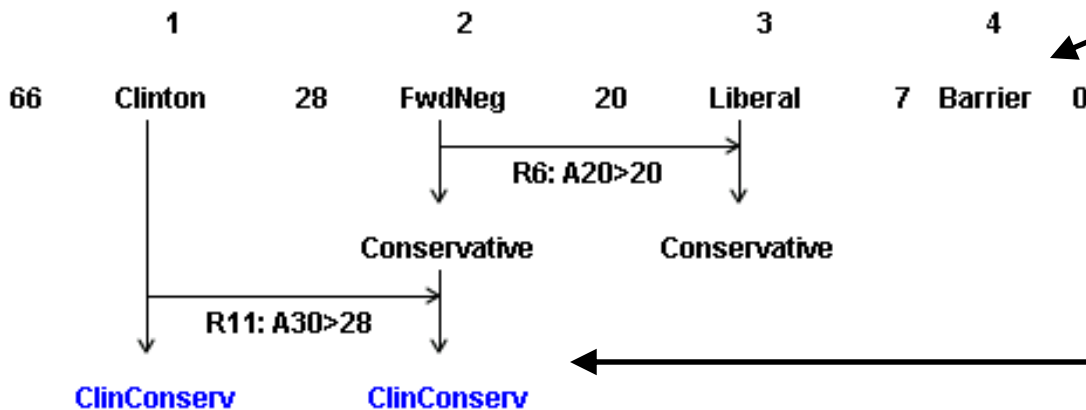
!2 (ge) by distancing himself somewhat from the sister soldier / jesse jackson issue , [1] he] really established himself as [2] not] being a traditional [3]liberal][4] .]

.....Scores this paragraph (P_ID=5486, P_NUM=2): ClinConserv=2

Paragraph text with dictionary words highlighted and numbered

Text equivalent with ideas in place of words and distances between ideas in number of characters

Rule usage e.g. R11 is rule number, A30>28 means rule meets distance and idea order constraint



ANES algorithm to code for knowledge about Rehnquist

- "We are strict regarding acceptable answers: We will accept ONLY 'Chief Justice' – 'Justice' alone is definitely *NOT* acceptable. (The court must be 'the Supreme Court' – 'Chief Justice of the Court' won't do. Note: applies only if R would specifically say 'the Court,' a rare phrasing, rather than 'the Supreme Court'). If unsure whether correct, code as best you can and record R's response as a remark."

InfoTrend format for Rehnquist algorithm from previous slide

```

----->---|<----->---|-----
Chief
 { chief } { head } { main }
Justice
 { honcho } { justice } { one }
Supreme
 { supreme }
Court
 { court }
ChiefJust      1 cSCORED      'Chief Justice'
SuprCourt      2 cSCORED      'Supreme Court'
 { united states }
ChiefJustUS    3 cSCORED      'Chief Justice of the United States'
----->---|<----->---|-----
FSR2           Acting:      No. spaces      R if operator
                Ahead      (-1 means anywhere)      retained      Equal
                Behind      |                          |          Priority
                Either      |          ADDIF1ST          DELBETWEEN    |      (E/e)
Nothing or      Direction  |          INSERT or          DELETED or    |      Rule
Operator----->      (A/B/E) |          <-Target----->      <-New target-> |      |
-----|-----|-----|-----|-----|
Chief           A          10      Justice      ChiefJust      1
Supreme         A          10      Court         SuprCourt      2
ChiefJust       E          10      SuprCourt     ChiefJustUS    3
-----|-----|-----|-----|-----

```

Crisis management

- At regular intervals in time and after crises, ask very general open ended questions
 - E.g. reasons for liking or disliking a political candidate
 - E.g. country's most important problem
 - Avoids the delay needed for designing good closed ended questions
- Score for unexpected ideas as they arise to explore the salience of the ideas over time
 - Useful for knowing if and when to react to adverse information
 - E.g. only respond to a new attack ad if its ideas are prominent in the open end responses
 - The InfoTrend system is well suited for the scoring because the algorithm can be written rapidly for any particular issue using logic mimicking human thought