
TEXT CLEANING WITH LEXOS



COMMON ISSUES WITH DIGITIZED HISTORICAL TEXT

- Newspaper column breaks
- Page structure (Table of Contents, Indexes, etc.)
- Condition of the original physical document
- Type of text contained within the document (handwritten v. font type)
- Year the original document was created
- Maturity of the OCR algorithm used at the time of creation
- Images

because of Alert Press, Congress

Ford Sees No Opportunities for New Watergate Scandal

WASHINGTON, Aug. 8 (AP)—President Ford said yesterday that "there would never be an opportunity for another Watergate to take place" because of an alert news media, public and Congress.

Richard Nixon announced his resignation as President a year ago today because of the Watergate scandal and Mr. Ford replaced him the following day, Aug. 9.

Mr. Ford said the strong U.S. system of government enabled the nation to survive that "very dramatic period."

"If you look at what happened before Aug. 9 and what happened since, I think you will come to the conclusion that the system did work," Mr. Ford said in a public Broadcasting Service television interview.

Reflecting on his first year in office, Mr. Ford said he felt the uniqueness of his administration had restored the confidence of the U.S. people in their government, that he had established a third presidency and that he was winning the fight against inflation and recession.

Mr. Ford said he and his family enjoyed the White House, which he called a "magnificent home, a super place to live." But he said "it can be a little lonely sometimes."

It also gives you an opportunity to focus in on the problems and it is the problems that come with the house that make it somewhat difficult at times to really relax and enjoy ourselves."

The President, reminded by an interviewer that many had pardoned him as an interim president when he took office, was asked if he had established his own presidency.

I think we have gradually put together a domestic program and a foreign policy that can be identified as a Ford administration," he replied. "As we move in the next 12 to 14 months, I think it will become more and more evident, which, of course, gives us an opportunity in the election to lay that record so the public can judge it

ment, for a successful implementation of economic policy that will provide jobs in the private sector in the long run and in foreign policy will expand our efforts for peace throughout the world," he said.

Mr. Ford dwelled on the Middle East during the interview and warned Israel and Egypt that they must be more flexible in their negotiations. He said there is a real possibility of confrontation between the United States and the Soviet Union if war breaks out again in the Middle East.

"Both Israel and Egypt have to understand that flexibility at this crucial time is important to the peace of that area and possibly peace in the world. Israel has to be more flexible and I think Egypt has to respond."

He also said that the House of Representatives has jeopardized U.S. foreign policy and undercut national security by refusing to lift the arms embargo against Turkey. He called the House refusal "the most serious, wrong decision since I have been in Washington, which is 27 years."



United Press International
Richard Nixon bids farewell to his White House staff on Aug. 9, 1974, the day after he announced his resignation. With him is his wife, Pat, and daughter Tricia . . .

A Year After Resigning, Nixon Remains a Lone

By Kenneth Reich

SAN CLEMENTE, Calif., Aug. 8.—A year ago today, Richard Nixon announced his resignation from the presidency at the climax of an unprecedented personal and governmental scandal.

The next day, he arrived here, beginning what his White House news secretary, Ronald Ziegler, has called "the first American political exile."

It remains by most evidence a sad, often lonely experience. Mr. Nixon seldom emerges into public view; his health remains precarious; his views toward what has befallen him often are bitter and doubt often is expressed by those in contact with him that his much-heralded book—his only current attempt at public self-expression—for which he stands to receive a \$2.5-million advance, will be a candid one.



"Ford Sees No Opportunities for New Watergate Scandal." International Herald Tribune [European Edition], August 9-10, 1975, p. 3. International Herald Tribune Historical Archive, 1887-2013

SAMPLE | SCAN

s ford Sees No Opportunities for New Watergate Scandal ■, WASHINGTON, Aug. 8
d never be an (opportunity for another Water(jSte to take place" because of K
ixon announced his (resignation as President a year ago today because of the
ng day, (Aug- 9. I Mr. Ford said the strong U.S. ■system of government enable
iod." I "If you look at what happened (before Aug. 9 and what happened I sinc
m I did work," Mr. Ford said in a (public Broadcasting Service teleI vision i
. Ford said he felt the ■openness of his administration I had restored the co
he had established a I Ford presidency and that he was I winning the fight a
his fam■ily enjoyed the White House, I which, he called a "magnificent I hom
little lonely [at times." ["It also gives you an opportunity to focus in on
ouse that make it somewhat difficult at times to really relax and enjoy I you
any had regarded him as an interim President when he took office, was asked i
gradually I put together a domestic program and a foreign policy that can j b
move in the next 12 to 14 months, I think it will become more and more eviden
xt election to lay that record out so the public can judge it against any alt
ed about criticism that he had not provided a broad map for the country's fut
programs because he wanted first to get the nation's fiscal situation under I
hink we have made substantial headway–then we can take a look at it and make
next State of the Union message will include some subjects that will point i
e known, in my judg-,ment, for a successful implementation of economic policy
run and in foreign policy will expand our efforts for peace throughout the w
the interview and warned Israel and Egypt that they must be more flexible in
of confrontation between the United States and the Soviet Union if war breaks
to understand that flexibility at this crucial time is important to the peac
el has to be more flexible and I think Egypt has to respond." He also said th
U.S. foreign policy and undercut national security by refusing to lift the arm
al "the most serious, wrong decision since I have been in Washington, which i
11 I i 1 II m I lli V ma i * S §1? til i I h ! m * ■ p >■ s < p m Sag M ■ : ■
mm llflfIPPt United Press International. Richard Nixon bids farewell to his
nounced his resignation. With him is his wife, Pat, and daughter Tricia . . .

SAMPLE | OCR

THE CONTENTS.

BOOK III.

Of WORDS.

CHAP.

- 1. Of Words or Language in general.
 - 2. Of the Signification of Words.
 - 3. Of General Terms.
 - 4. Of the Names of simple Ideas.
 - 5. Of the Names of mixed Modes
- and Relations.
-

CHAP.

- 6. Of the Names of Substances.
- 7. Of abstract and concrete Terms.
- 8. Of the Imperfection of Words.
- 9. Of the Abuse of Words.
- 10. Of the Remedies of the foregoing Imperfections and Abuses,

BOOK IV.

Of KNOWLEDGE and OPINION.

CHAP.

- 1. Of Knowledge in General.
 - 2. Of the Degrees of our Knowledge.
 - 3. Of the extent of Humane Knowledge.
 - 4. Of the Reality of our Knowledge.
 - 5. Of Truth in General.
 - 6. Of universal Propositions, their Truth and Certainty.
 - 7. Of Maxims.
 - 8. Of trifling Propositions.
 - 9. Of our Knowledge of Existence.
-

CHAP.

- 12. Of the Improvement of our Knowledge.
- 13. Some other Considerations concerning our Knowledge.
- 14. Of Judgment.
- 15. Of Probability.
- 16. Of the Degrees of Assent.
- 17. Of Reason.
- 18. Of Faith and Reason, as contra distinguished.
- 19. Of Enthusiasm.

SAMPLE 2 SCAN

Locke, John. An essay concerning human understanding. In four books. Written by John Locke, Gent. ... 10th ed., vol. 2, printed for Edward Symon, over against the Royal Exchange in Cornhill, M.DCC.XXXI. [1731].

AN E S S AY, CO NCERNING Human Underfianding.,.BY O H AL O C K E, Gent. VOLUME II., p L O N D O N: Printed for EDWARD S Y M O N,
over, against the Royal Exchange in Cornhill, M.DCC.XXXI /[, __, - !, I -----
THE CONTENTS., BOOK III, OF WORDS, CHAP. CHAP. x. Of Words or Language in ge- 6. of the Names of Sufli.aces., neral 7. of abfltrat mand
concrete Terms. 2. of the Signification of Words. 8. of the Imperfession of Words. 3. Of General Terms. 9. of the Abuse of Words. 4. of
the Names of /imple Ideas. 0. Of the Remedies of the forega- 5. of the Names of mixed Modes ing Imnperfeffions and AbrJfes,, and
Relations., BOOK IV. OF KNOWLEDGE and OPINION, CHAP. j. Of Knowledge in General. 2. of the Degrees of ourKnowledge. 3. Of the extent if
Humane Know-, ledge. 4. Of the Reality of our Knowledge. 5. Of Truth in General. 6. Of universal Propositions their, Truth and Certainty.
7. Of Maxim., 8. Of trifling Propositions. 9. Of our Knowledge of Exience. 10. Of the Exflnce of a GO D. 1. Of the Knowledge of the
Exif-, tne of oth,er Thingr., /3 h, CH A P. 12. Of the Imnproement of our, Inomledge. 13. Some other Considerationt con-, cerning our
Knowledge. 14. Of udgment. 15. Of Probability. 16. Of the Degrees of A4ent 17. Of Reaon. 8. Of Faith and Reaffbon asr on., tra
difflngguilhed. 19, Of Enthuij7fir. 20. f wrong n.4set or Error. 21. The Division of the Sciences. A r THEA

SAMPLE 2 OCR

of the matter in Mackworth's favor —
He has conducted himself with
much firmness & composure during
this delay which he feels to have been
unavoidable & which he considers
in its just light —

D. M. is today a great liberty in
thus intruding himself upon Mr. Peel's
privacy & had he been in London
intended to have waited upon him
personally — D. M. feels sure that
Mr. Peel will forgive this intrusion
to which D. M. has been a
little emboldened by the recollection
of old times at Hanover & at Oxford
where Mr. Peel has so understandingly
borne in mind from time to time —

33. Bedford Place
Baptist Chapel
Friday
August. 12. 1825.

SAMPLE 3 SCAN

Petitions Dh-Dm; Anton MacKenroth (Anthony MacKenna). 1825
May - 1825 Nov. MS Crime and the Criminal Justice System:
Records
from The U.K. National Archives: HO 17 Records created or
inherited by the Home Office, Ministry of Home Security, and
related bodies, Home. Office: Criminal Petitions, Series I HO
17/30/109. The National Archives (Kew, United Kingdom).

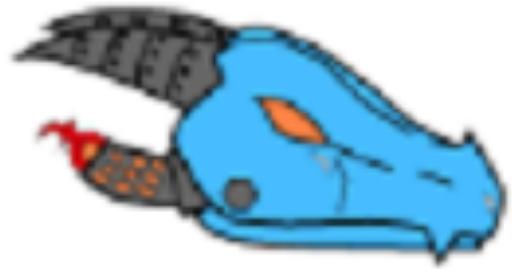
FOR PUBLISHING BV SUBSCRIPTION ^{to} TREATISE ON BOOK-KEEPING, APPLICABLE TO THE BREWERY., ¹⁸¹¹ By EDWARD N.
HAYMAN, COMMON BREWER, Author of a Practical Treatise on Brewing, which has passed through Three Editions, THE success which the Author
has met with by the publication of his "Treatise on Brewing," has induced him to submit his System of Book-keeping., as applicable to
the Brewery, to the inspection of the Trade ; but although the great importance of a correct System of Accounts, in a business like
the Brewery, which is exposed to such vast peculation from the numerous parties employed therein, is acknowledged by every person
connected with it, still the Author is aware of the questions that may arise as to the utility of any new work on the subject ;
however, as there is not one work that is exclusively applicable to the Brewery, the Author is flattered by hopes that his present
undertaking may meet with the same liberal reception which his former work has experienced. The present work is founded on the
infallible principle of double entry, as well as by single entry, and by a peculiar arrangement of the various accounts, possesses the
following advantages over the general method adopted in most Breweries. First. It embraces the power of detecting any wilful
expenditure or misapplication of the various articles employed in the Brewery, by the assistance of one or two auxiliary books.
Secondly. The Journal becomes not only a correct check upon the accuracy of the Ledger, but also upon the auxiliary books employed.
Thirdly. The fluctuation of price in the several articles used in the Brewery is clearly exhibited. Lastly. The present system will
point out in the short space of one hour the actual net profit per barrel, however extensive the trade of the Brewery, without the
trouble of taking- stock, or striking a balance himself a fair living profit; and at the same time he will be
able to discover in what particular department any exorbitant or wilful expenditure takes place, and thereby to strike at the root of
the evil at once. The Work is founded not on chimerical ideas, but on the actual transactions of one of the most extensive concerns
of ^{any} day as an Ale and Table Beer Brewery in London, and by founding the Work on bonafide transactions, the correctness of the
system is discerned by the method of ascertaining the profit per barrel tallying with the balance as struck at the end of the year,
or. more properly speaking, with the profit at the close of the year as appears by the balance sheet. With these various advantages
the Author looks with some hope to the success of his present undertaking, and should it vie with his - Practical Treatise on
Brewing," his labour will be fully repaid. The Work will be published in Octavo, by the same respectable Firm as his former one, i.e. Longman,
Hurst, Rees, Orme, and Co., Paternoster Row; and will be put up as such a number of Subscribers, all, have been obtained
as will cover the expenses of publication and printing.

SAMPLE 3 OCR

WHAT IS LEXOS?

From the [Wheaton College Lexomic Department Website](#):

“Lexos is a web-based tool to help you explore your favorite corpus of digitized texts. Our primary motivation is to help you find the explorer spirit as you apply computational and statistical probes to your favorite collection of texts. Lexos provides a workflow of effective practices so you are mindful of the many decisions made in your experimental methods.”



{L}exos

[HTTP://LEXOS.WHEATONCOLLEGE.EDU/UPLOAD](http://LEXOS.WHEATONCOLLEGE.EDU/UPLOAD)



Upload

Browse...

?

Scrape

Scrape

?

Drag Files Here

Enter URLs to scrape here.

Upload List

?



Upload

Browse...

?

Scrape

Scrape

?

Enter URLs to scrape here.

Drag Files Here

Upload List

?

Nixon_denies_Watergate_IO0702817044.txt



Active	#	Document	Class	Source	Excerpt	Download	?
●	1	Nixon_denies_Waterga_IO 0702817044		Nixon_denies_Waterga_IO 0702817044.txt	Nixon denies Watergate cover-upBy STEPHEN BARBER Washington /CLAIMS Mr John Dean, White House ^ legal adviser dismissed week, P... ...e House gaps left week's resignations dismissals. Pentagon Papers trial, Watergate developments, & picture—P4; Peterborough—P18	Download	?

**Scrubbing Options**

- Make Lowercase Scrub Tags [Options](#) [?](#)
- Remove Digits Remove Punctuation
- Remove Spaces Keep Hyphens [?](#)
- Remove Tabs Keep Apostrophes [?](#)
- Remove Newlines Keep Ampersands [?](#)

Lemmas[Upload](#)[?](#)

Enter lemmas here.

Consolidations[Upload](#)[?](#)

Enter consolidations here.

Stop/Keep Words[Upload](#)[?](#)

- Off Stop Keep

Enter stop words and keep words here.

Previews[Preview](#)[Apply](#)[Download](#)[Nixon_denies_Watergate_IO0702817044](#)

Nixon denies Watergate cover-upBy STEPHEN BARBER
Washington /CLAIMS Mr John Dean, White House ^ legal
adviser dismissed week, P... e House gaps left week's
resignations dismissals. Pentagon Papers trial, Watergate
developments, & picture—P4; Peterborough—P18

Special Characters[Upload](#)[?](#)

- None MUFI 3
- Early English HTML MUFI 4
- Old English SGML

Enter special characters here.

[Page Walkthrough](#)[Page Help](#)[Help Glossary](#)[About Lexos](#)

The Scrubber Tool

Pre-processing your texts, what we refer to as "scrubbing", is a critical step when probing digitized texts and is a primary step in the Lexos workflow. In order to facilitate a conscious consideration of the many small decisions required, scrubbing options are isolated into individual choices. If for no other reason, your careful deliberation and choice of the many options during scrubbing facilitates replication of your analyses in the future, both by you and others who wish to replicate your

Scrubbing Options

- | | | | |
|--|--|-------------------------|-------------------|
| <input checked="" type="checkbox"/> Make Lowercase | <input type="checkbox"/> Scrub Tags | Options | ? |
| <input checked="" type="checkbox"/> Remove Digits | <input checked="" type="checkbox"/> Remove Punctuation | | |
| <input type="checkbox"/> Remove Spaces | <input type="checkbox"/> Keep Hyphens | ? | |
| <input type="checkbox"/> Remove Tabs | <input type="checkbox"/> Keep Apostrophes | ? | |
| <input type="checkbox"/> Remove Newlines | <input type="checkbox"/> Keep Ampersands | ? | |

Lemmas

[Upload](#)[?](#)

Enter lemmas here.

Stop/Keep Words

- Off Sto

Enter stop words a

Special Characters

- None
 Early English H
 Old English SG



Scrubbing Options

- Make Lowercase Scrub Tags [Options](#) ?
- Remove Digits Remove Punctuation
- Remove Spaces Keep Hyphens ?
- Remove Tabs Keep Apostrophes ?
- Remove Newlines Keep Ampersands ?

Lemmas

[Upload](#) ?

Enter lemmas here.

Consolidations

[Upload](#) ?

Enter consolidations here.

Stop/Keep Words

[Remove Upload](#) ?

Upload: English_Stopword_List.txt

 Off Stop Keep

Enter stop words and keep words here

Previews

[Preview](#)[Apply](#)[Download](#)

Nixon_denies_Watergate_IO0702817044

Nixon denies Watergate coverupBy STEPHEN BARBER
Washington CLAIMS Mr John Dean White House legal

Scrubbing affects all active documents and cannot be undone. So make sure to deactivate any documents you do not wish to scrub using the [Manage](#) tool. If you apply scrubbing and later wish to revert to the unscrubbed version, you will have to upload another copy to Lexos.

Special Characters

[Upload](#) None MUFI Early English HTML MUFI 4 Old English SGML

Enter special characters here.

LEARNING MILESTONE: IDENTIFYING ISSUES

OCR

- SAMPLE OCR 4:
 - Newspaper column breaks
 - Page structure (Table of Contents, Indexes, etc.)
 - Condition of the original physical document
 - Type of text contained within the document (handwritten v. font type)
 - Year the original document was created
 - Maturity of the OCR algorithm used at the time of creation
 - Images
- SAMPLE OCR 5:
- SAMPLE OCR 6:



SAMPLE 4 SCAN

"Seagers Gin." Sunday Times, 15 Dec. 1940, p. 2. The Sunday Times Historical Archive

, . - i, s E Gin makes the Cocktail – gEAGE RS make the gin CEAGERV GIN liiiiiN'Hiiiinar ^Siiiiuiuiiuixa M 01S1IUEBI Of fINE
GIN'ttd&fOO OVEI IX> VEARS

SAMPLE 4 OCR

be there as at the cities, & A writes it this
"Morty" is wt rec'd p. J. where fort to be recorded.

Aug 18/58, H. writes to Morty. Not rec'd p. J.

Aug 26/58, H. writes to Morty. Rec'd p. J. & recd.
Sept. 17/58, T. R. S. "remedies", gives L.L. & W.D.
Richards a rec^t for her note of Sept. 7, 4 mos., \$571.97,
& agrees to hold by Bloomer accountable for
it until (an^o note) - (\$100. due but
prior of this because until after trade made
& attachment) And Richards here a
receipt, I do prod, for same debt - But
~~Bloomer has lost his note to T. R. S.~~
As

Attached \$100. Due 3rd, & letter of
L. Elmore Due 4th. None of his copy.
J. R. S. to L. Elmore - & in con. wth T. R. S., he
has got his note to him.

SAMPLE 5 SCAN

Dana, Richard Henry. Tremlett vs. Tremlett Sale of Ship—Brig
Bloomer, Case no. 1819. October 16, 1858-January 8,
1859. MS Nineteenth Century Crime: Manuscripts and Ephemera from
the American Antiquarian Society: Dana,
Richard Henry, Jr., Legal Papers, 1729-1895 Case no. 1819. American
Antiquarian Society

£ I) c o m m o n w c a I i I) of ill a i β a 1 1) u s c t i i . Suffolk, ss, To the Sheriff of our County of Suffolk, or his Deputy, GREETING. WE COMMAND you to attach the Goods or Estate of mm/A (f:\ tC ^;fafa: Д^^,Д^ to the value of Н^я. A ' /;tyr/L-- Dollars, and summon the said I t Lt/~/ ^ ^ e та У De fº un d w yº ur precinct, to appeat- before Our Justices of Our Supreme Judicial Court, next to bo hokien at Boston, within and for Our County of Suffolk, on the y\лA t" Tuesday of ' ^£. next ; then and there in Our said Court, to answer unto Яр л. ~* t л w^ OIA - c tL (A^cL t%LCJuu;CA И- daaruit Crf };■ 'A /уъ&y ;L rtv,Ac 'л^-^,Ir ' s r To the damage of the said р/U /кнA и T^sLeL&JU/L· (as ^Ä« t, .w//) the Sum of ft01V1 \X^ЛлAШлЛ* Dollars, which shall t^en and there be made to ap- pear, with other due Damages. And whereas the said J uh-iyIt u~on say , that the said 'ti- 1 ^'^ f. pJiAi-y^t/AA . _ ha t> -not in κ^</ own hands and possession, Goods and Estate to the Value of ' £j0 7 /ú t \^ / ; ^? Dollars aforesaid, which can be come at to be attached, but has intrusted to, and de- posited in, the hands and possession of J^h^ /. f^uyUFU c Y /' ууь y t /t v ý.A:n /: d^jt^L.*^, \/ýry^f / АыЛ^*р^ / to^ A^елС^А <y~rr_ χι&A A. ftiü. (Y -tL èd/iroui J. ишмлХку Ia/uma» A-'/1-y^/ / /yAyy> % . cjJL ^.^v - T ,. Mtiec 7 0f lllic said ^5Z?BMГ , 10 ods, . Effecte,,and Credits to tho said Value: WE COMMAND you, therefore, that you summon the said,'/Altt^A,, if Mie k may be found in your precinct, to appear before Our Justices of Our said Court to be/hoklen as aforesaid, to shew cause if any /het/ lictft\v\\v Execution,to be issued upon such Judanenfras the said , { /y i ; /y //ý / may recover,against the said /ÝIA/yL i- Al f ' y y/lt-^r w ;is Action, (if any; should not issue against />£<7 Goods, Effects, or Credits, in the hands and possession of//<i^ the said And have you there this Writ, with your doings therein. Witness, LEMUEL SHAW, Esquire, at Boston, the - UHuJrt day of oo-fff { y f L^_ in the year of our Lord one thousand eight hundred and (mj fqu fr-Jfc^ /'it . &■ AA/AA^ / Clerk.

SAMPLE 5 OCR

~~Mrs. Benson~~
~~Dr. Matsner~~

October 14, 1935
e

In talking with Mrs. Huse this morning,
she asked if you had seen the appointments
to the committee to investigate contraception.
Said you would be interested in the affiliations
of the individual men:

McCord--strong for b. c.
Kosmak of the National Committee--lukewarm
Woodward--teaches in Georgetown--an R. C.
O'Shea--R. C.
Carl Davis--Marquette Univ. --R. C.
John Rock--she says is for b. c.

g c

Page 1195

Holden, Frederick C., Eric M. Matsner, and Olin West. American Medical Association, Committee to Study Contraceptive Practices and Allied Problems, 1935-1937. 1935-May, 1940. MS
Planned Parenthood Federation of America Records, 1918-1974: Part 2: Correspondence and Subject Files, 1920-1974: Subject Files. Sophia Smith Collection, Smith College Library

SAMPLE 6 SCAN

Octoter 14, 1935 In talking with Mrs. Huse this morning, she asked if 70 a had seen the appointments to the committee to inyestigge
contraception. Said you would he interested in the affilia« tions of the intiåådoal men; McCord-strong for h. c. Eo smak o f t he Nat
ional Görnit tee-lukewam Woodward-teaches in Georgetown-an R, C. O'Shea- B , C. Carl Davis-Marquette Uniy. - B. C 1 John Bock-she says
is for h. c. m g

SAMPLE 6 OCR