

Appendices

This document contains supporting information for the article “Towards a scientific workflow featuring Natural Language Processing for the digitisation of natural history collections (Owen, D. et al. 2020)”.

1 OCR Software Settings	2
1.1 Tesseract 4.0.0 and Tesseract 3.0.51	2
1.2 ABBYY Finereader Engine 12	2
2 OCR Line Correctness Analysis	3
2.1 Line Correctness with Whole Images	3
2.2 Line Correctness with Segmented Images	32
3 Named Entity Recognition Analysis	73
3.1 Named Entity Gold Standards	73
3.2 Stanford NER using OCR text from Whole Images	76
3.3 Stanford NER using OCR text from Segmented Images	79
4 Non-standard Terminology Extraction Analysis	83
4.1 FlexiTerm Output Using OCR Text From Segmented Images	83

1 OCR Software Settings

Settings are provided for the OCR programs used in section 3. This is to aid reproducibility of results.

1.1 Tesseract 4.0.0 and Tesseract 3.0.51

Page segmentation mode: "3" (Fully automatic page segmentation)

To not create accompanying xml file of results

- tesseract_create_hocr "0"

To remove noise in the input image

- textord_max_noise_size "45"

To specify that the space between word is variable

- textord_space_size_is_variable "1"

To not run in parallel

- tesseract_parallelize "0"

To load Directed Acyclic Word Graphs

- load_system_dawg "0"

1.2 ABBYY Finereader Engine 12

Load predefined profile: "TextExtraction_Accuracy"

Plain text export format

- FileExportFormatEnum.FEF_TextUnicodeDefaults

Rich text export format

- FileExportFormatEnum.FEF_RTF

2 OCR Line Correctness Analysis

Column headings denote image file name and number of printed lines contained within. The bracketed number in each column denotes the number of lines of printed text. For example, “B 10 0002520 (/24)” concerns an image named “B 10 0002520” containing 24 lines of printed text.

Segmented image naming convention

IMAGENAME+TYPE+NUMBER

IMAGENAME = the original file from which the fragment was extracted

TYPE = The type of segment detected (lbl=label, bcd=barcode, clc=colour chart)

NUMBER = sequential item number within herbarium sheets

Example:

L.2168354_lbl02.jpg

IMAGENAME = L.2168354_lbl02.jpg (original file: L.2168354.jpg)

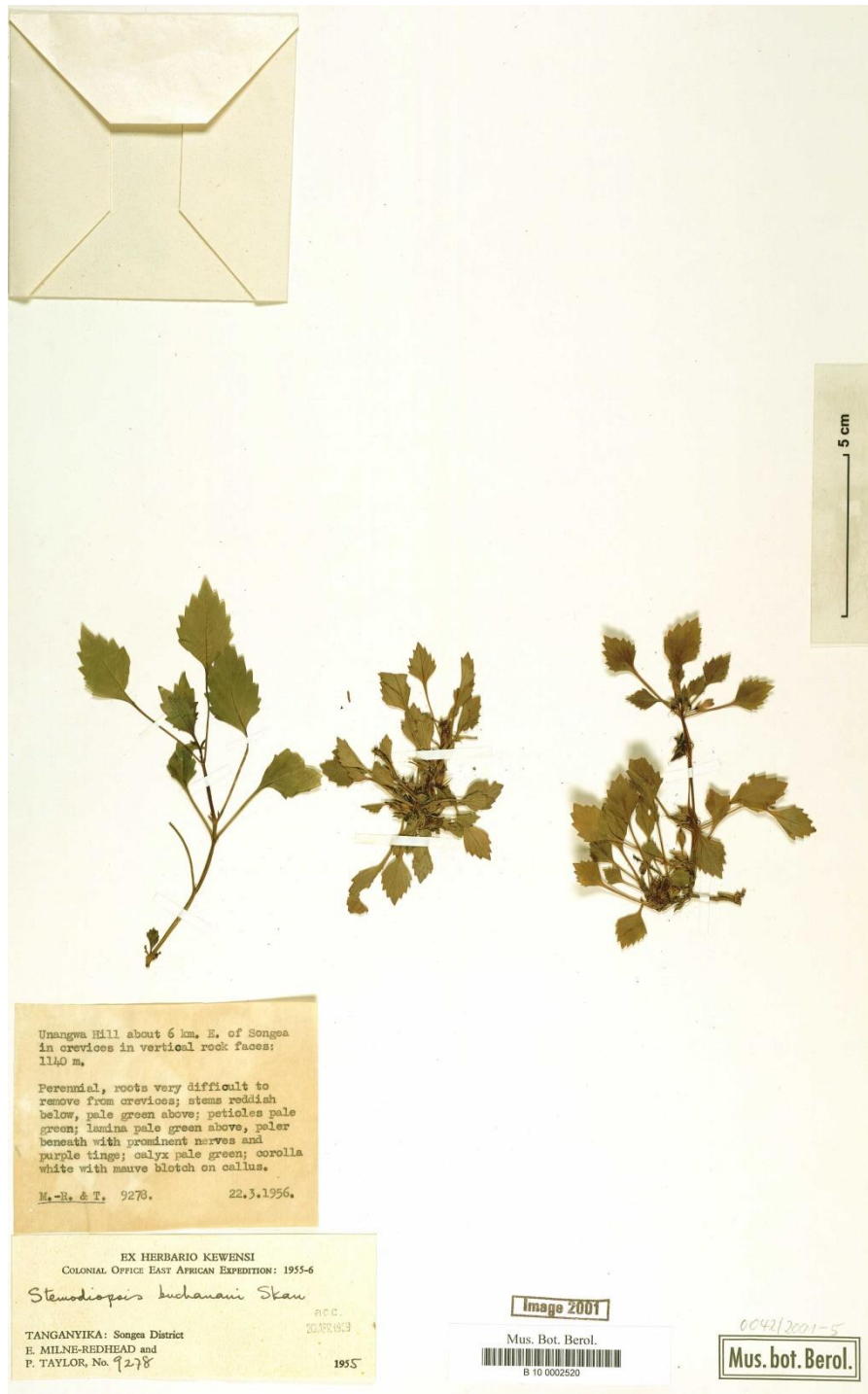
TYPE = lbl (label)

Number = 2

2.1 Line Correctness with Whole Images

Whole images	B 10 0002520 (/24)	BM000500117 (/14)	E00015443 (/10)	FIG.2770 (/15)	K000025814 (/32)
Tesseract 4.0.0	12	10.5	7.5	10.5	30
Tesseract 3.0.51	9	7	3	4	24.5
ABBYY FineReader Engine 12	19.5	3.5	9	10.5	30
Microsoft OneNote 2013	15	10.5	7.5	12.5	31.5

2.1.1 Image "B 10 0002520"



Manual transcription of **B 10 0002520** | Number of lines: 24

Mus. Bot. Berol.

B 10 0002520

5cm

EX HERBARIUM KEWENSI

COLONIAL OFFICE EAST AFRICAN EXPEDITION: 1955-6

TANGANYIKA: Songea District
E. MILNE-REDHEAD and
P.TAYLOR, No. 9278
1955
acc.
20.APR.1959
Unangwa Hill about 6 km. E. of Songea
in crevices in vertical rock faces:
1140 m.
Perennial, roots very difficult to
remove from crevices; stems reddish
below, pale green above; petioles pale
green; lamina pale green above, paler
beneath with prominent nerves and
purple tinge; calyx pale green; corolla
white with mauve blotch on callus.
M.-R. & T. 9278. 22.3.1956.
Image 2001
Mus. bot. Berol.

Tesseract 4.0.0 | B 10 0002520 | Score: 12/24

|
|
| ;
|
|
|
CL t—
Unangwa Hill about 6 = Ee of Songea
in crevices in verilesl. rock Faosg; :
1140 m,
Perennial, roots very aiffiult to
remove from crevices; stems redd:
below, pale green eve; Poth
green; Lauila pale green a
EX HERBARIO KEWENSI
CoLoNIAL OFFICE EAST AFRICAN EXPEDITION: 1955-6
a0
TANGANYIKA: Songea District

E. MILNE-REDHEAD and
| P. TAYLOR, No. 1 2%
of 1004509
GANT 13
1955”
A
Mus. Bot. Berol.
HL
15cm
Mus. bot. Berol.

Tesseract 3.0.51 | **B 10 0002520** | Score: 9/24

\
\
\
i
v
\
M
13mm Hill ahead: 6 km. E of Song”
11 Meet! in Wind rock fauna;
1m 3.
Perennial, roots very Wm”
rm: rm ormimm; atom
balm, palm: yam mum; m1
wan; lamina mum
EX HERBARIO KEWENSI
COLONIAL OFFICE EAST AFRICAN Expnmon 1955 6
mm”
TANGANYIKA Songea District
E MILNE REDHEAD and
P TAYLOR No ‘12,}?
r“1}
Lv’ M.
1955’
'i
Mus Bot Berol
HIIHJmllliiHIIIMILIIQMMQIMIIIIIIIIIIIIWIH
IScm

(
Mus bot Berol

ABBYY FineReader Engine 12 | B 10 0002520 | Score: 19.5/24

I
Perennii
purple tinge; calyx pale green; corolla
white with mauve blotch on callus.
22.3.1956.
ku-cX^a-^_a-v<Jv <SV<^vv
Hmafle fOOf |
ocw(k
1
Mus. Bot. Berol.
Mus. bot. Berol.
HniRIIIIIIIWIIIIIRIKIIIIHnH
195.S-
B 10 0002520
§
m
EX HERBARIO KEWENSI
Colonial Office East African Expedition: 1955-6
Unangwa Hill about 6 lon. E. of Songea
in crevices in vertical rock faces:
1140 m.
TANGANYIKA: Songea District
E. MILNE-REDHEAD and
P. TAYLOR, No.
_____, roots very difficult to
remove from crevices; stems reddish
petioles pale
„ above, paler
sneath with prominent nerves and
acc.
2GMi9L9
J®1?1
JS*"
Lal,
Fror

below, pale green above;
green; lamina pale green
beneath with prominent **nt**

Microsoft OneNote | **B 10 0002520** | Score: 15/24

Unangwa **Eli 11** about **6** E. of Songea
in crevices in vertical rock faces:
Perennial 9 roots very difficult to
remove from crevices; **sterns** reddish
below, pale green above; petioles pale
green; **la.rl)ina** pale green above, paler
beneath with prominent nerves and
purple tinge; calyx pale green; corolla
white with mauve blotch on callus.

& TO 927B.

22.3.1956,

EX HERBARIO KEWENSI

COLONIAL OFFICE EAST AFRICAN EXPEDITION: 1955-6

TANGANYIKA: Songea District

E. MILNE-REDHEAD and

P. TAYLOR, No.

ma e 2

Mus. Bot. Berol.

B 10 0002520

E

10

MUS. **boto** Berol.

2.1.2 Image “BM000500117”



Manual transcription of **BM000500117** | Number of lines: 14

THE NATURAL HISTORY MUSEUM, LONDON
DEPARTMENT OF BOTANY
000500117 (BM)
Vidi S. Pérez-Ortega
5th December 2010
Herbarium Musei Britannici
Parmelia perlata (Huds.) Ach.
TASMANIA: Tinderbox (near Hobart).
Alt. sea level+.
Common on dolerite in open dry sclerophyll
Forest.
G.Kantvilas No. 277/80 20.7.1980
Det: P.W. James TLC: stictic complex
Determinavit

Tesseract 4.0.0 | **BM000500117** | Score: 10.5/14

Herbarium Musei Britannici
Parmelia perlata (Huds.) Ach.
TASMANIA: Tinderbox (near Hobart).
Alt. sea level+.
Common on dolerite in open dry sclerophyll

forest.
G. Kantvilas No. 277/80 20.7.1980
Det: P.W. James TLC: stictic complex
THE NATURAL HISTORY MUSEUM, LONDON
Tx cep lu oA 4 DEPARTMENT OF BOTANY
| |
wr im
\$ leuwdoadputt
Determinavit
10 Lye
; TEN
copyright reserved MUSEUM
5s
213
ES
| 2%
3%
S| \$
2 2
4
Sis
= |S

Tesseract 3.0.51 | **BM000500117** | Score: 7/14

Herbarium Musei Britannici
Parmella Eclata (hm) Acb
TASMANIA Tinderbox (near hobart)
Alt sea level+
Common on dolerite 10 **open dry** sclelopnvl
forest
G Kantvilas **No 277/80** 20 71380
Det P W James TLC **stlctlc** complex
THE NATURAL HISTORY MUSEUM LONDON
W \ "IL IL L n i DEPARTMENTOFBDTANV
] '
72:" 7 ,7 HlmmwmmwW
< Wilcdzgx [;
NLHVHHUd
10 NATURAL

HISTORY
copyright reserved
MUSEUM
ma
B'N
*s.
5H,.
:73
b E
&' §
['3
§\
>1u

ABBYY FineReader Engine 12 | **BM000500117** | Score: 3.5/14

Herbarium Musei Britannici
Parmelia perlata (Huds.) Ach.
(n
Common on **dolerito** In
tXi:
kjLU^XitL
"7^.. (w-
MOLOWt
\$00((7
I
8 910 r^n history
jht reserved
U MUSEUM
gF'
<<
■ ■ ■
■ ■
(near Hobart).
n open dry sclerophyll
M
7 8
copyright reserved
h
il

jf

Microsoft OneNote 2013 | **BM000500117** | Score: 10/14

Herbarium Musei Britannici

Parmelia **oerlata** (Huds) Ach.

TASMANIA: Tinderbox (near Hobart) .

Alt. sea level+.

Common on dolerite in open dry sclerophyll forest.

G. Kantvilas No. 277/80 20.7.1980

Det: P.W. James TLC: stictic complex

Sic . m

THE NATURAL HISTORY MUSEUM, LONDON

DEPARTMENT OF BOTANY

0005001 i i (8M)

Determinavit

2.1.3 Image "E00015443"



Manual transcription of **E00015443** | Number of lines: 10

ROYAL BOTANIC GARDEN
EDINBURGH
E00015443
HERB. HORT. EDINB.
Flora from Kunming Institute, China
Corydalis
Fumariaceae
B & L 12,228
Gang-Ho Ba, Lijiang
September 1987

Tesseract 4.0.0 | **E00015443** | Score: 7.5/10

HERB. HORT. EDINB.
TTT TET
dh
copyright reserved
ROYAL BOTANIC GARDEN
yg
Flora from Kunming Institute, China
" & a }]
Corydalis sp Smithiana Rede 4,
Fumariaceae gid
B&L 12228
' y Gang-Ho Ba, Liii
- September 1987 ang

Tesseract 3.0.51 | **E00015443** | Score: 3/10

HERB HORT EDINB
l i l l |
IL" IHH Hil H
\umuumfififilfiifwumum
Flora from Kunming Institute, China
a a) N i '
Calydalzs sp 5 m ,Uw. a my RM; 1
Fumarmceae { Ly
B&L12228

' Gang H0 Ba, L
j September 1987 lllang

ABBYY FineReader Engine 12 | E00015443 | Score: 9/10

HERB. HORT. EDINB.

V —

>

9

A

V

1 "*"

|**

\

◆-'

|

niiTnnT

/

/

i

A

s

r>

:

'A

K

«

||

E00015443

//&

/ tn ,7Afa « ReA£L

Kc

Gang-Ho Ba, Lijiang

September 1987

LULU

r

||||||||||||

CD

CO

:
J
E
o o
'A
4
H
ROYAL BOTANIC GARDEN
EDINBURGH
l
r
1
i
■
■
■HL
: Y|
L
Flora from Kunming Institute, China
Corydalis sp
Fumariaceae
B & L 12,228
!
i
l
/A
£
-nrof
CO 2
O)
-l
o
o
21
r
■
-
<2"l
sits
773ft

- .
I/

Microsoft OneNote 2013 | **E00015443** | Score: 7.5/10

HERB. **HORTO** EDINB.

Luu-I

ROYAL BOTANIC GARDEN

EDINBURGH

E0001 5443

Flora from Kunming Institute, China

Col)'dalis sp-

Fumariaceae

B & L 12,228

September 1987

Gang-Ho Ba, Lijiang

2.1.4 Image "EIG.2770"



Manual transcription of **EIG.2770** | Number of lines: 15

Digitarium
<http://id.luomus.fi/>
EIG.2770
2013-04-02
MUSEUM BOTANICUM UNIVERSITATIS, HELSINKI
AUSTRALIA, Queensland, Bellenden Ker
National Park W. of Babinda.
The Boulders. Complex mesophyll rain
Forest.
10. Sept. 1981 Ilkka Kukkonen 10879
UMT grid:
MUSEUM BOTANICUM
UNIV. (H). HELSINKI
1459257
OUABR.

Tesseract 4.0.0 | **EIG.2770** | Score: 10.5/15

MUSEUM BOTANICUM UNIVERSITATIS, HELSINKI
Diplaziunn dilatatem 5.
AUSTRALIA, Queensland, Bellenden Ker
MUSEUM BOTANICUM
National Park W. of Babinda.
UNIV. {H}. HELSINKI is Fa
Complex mesophyll rain
ores .
um
EE [http://id.luomus. Fil](http://id.luomus.fi/)
145925
L270 EIG.2770
10. Sept. 1981 Ilkka Kukkonen 10879
[m]CTs. 2013-04-02
QUASR.
tar
i.
igi
D
UTM grid:

Tesseract 3.0.51 | **EIG.2770** | Score: 4/15

MUSEUM BOTANICUM UNIVERSITATIS HELSINKI
B'p'ul'um d""\\"£\rurn DC
AUEJLHQ'U LS, kglloenslgmd, 1h lленаen 'el
MUSEUM BOTANICUM Aatlondl idllx u. of 3001:1511
UNIV NH HLLSINKI #119
Hcgulders DONDMR mesopiy'l rain
ores .
lum
El'fiEi htth/idluomusfi/ 11;,925'7
J- r'l' , E|G2770 iU.)eit. 1051 1111c: u} omen 108 0
E15 4 2013 04 02 OUABR
tar
' \
lgl
D
UTMgrid 01W J 3M \\1U 1 '32

ABBYY FineReader Engine 12 | **EIG.2770** | Score: 10.5/15

T
in
o-

\
1 -
1
2

3

4

5
\
m

6

7

8

9

10

11

\\^

12

13

kV

<4/

l

r

o

1

W/

2

3

4

5

6

7

MUSEUM BOTANICUM UNIVERSITATIS, HELSINKI

8

O i p l t x Z. (U r v x J i l e x t t U R b ,

9

10

forest.

11

10. Sept. 1981

12

OUASR.

2013-04-02

UTM grid:

13

i

II

v

Ilkka Kukkonen IO879

det. J. S\rvela

1

AUSTRALIA, Queensland, Bellenden Ker
National Park W. of Babinda.

The Boulders. Complex mesophyll rain

MUSEUM BOTANICUM

UNIV. (H). HELSINKI

1459257

I

1

®!

I HS|B <http://id.luomus.fi/>

jffiH BG2''

yl

1

Microsoft OneNote 2013 | **EIG.2770** | Score: 12.5/15

10

11

12

13

10

11

12

13

MUSEUM BOTANICUM UNIVERSITATIS, HELSINKI

AUSTRALIA, Queensland, Bellenden Ker

National Park W. of Babinda

<http://id.luomus.fi/>

EIG.2770

2013-04-02

MUSEUM BOTANICUM

H E L S I N K I

UNIV.

1459237

OUAbr.

10.

UTM

The Boulders

forest.

Septe 1981

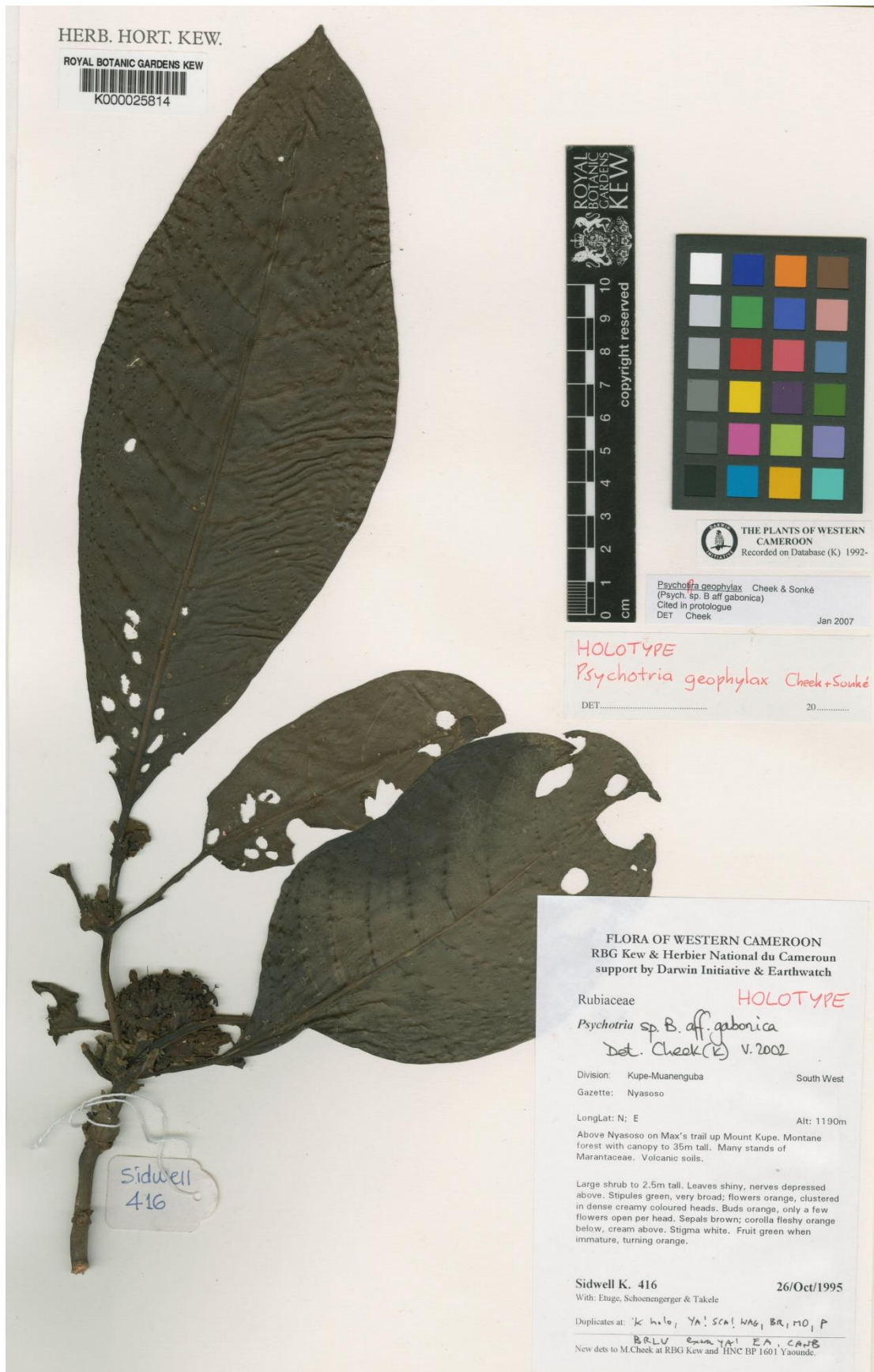
grid:

Complex mesophyll rain

Ilkka 10879

Jet, J. s

2.1.5 Image "K000025814"



Manual transcription of **K000025814** | Number of lines: 32

ROYAL BOTANIC GARDENS KEW

K000025814

THE PLANTS OF WESTERN

CAMEROON

Recorded on Database (K) 1992-

Psychotira geophylax Cheek & Sonké

(*Psych. Sp. B aff gabonica*)

Cited in protologue

DET Cheek Jan 2007

Det... 20...

HERB. HORT. KEW.

FLORA OF WESTERN CAMEROON

RBG Kew & Herbar National du Cameroun

support by Darwin Initiative & Earthwatch

Rubiaceae

Psychotria

Division: Kupe-Muanenguba South West

Gazette: Nyasoso

LongLat: N; E Alt: 1190m

Above Nyasoso on Max's trail up Mount Kupe. Montane

forest with canopy to 35cm tall. Many stands of

Marantaceae. Volcanic soils.

Large shrub to 2.5cm tall. Leaves shiny, nerves depressed above. Stipules green, very broad; flowers orange, clustered in dense creamy coloured heads. Buds orange, only a few flowers open per head. Sepals brown; corolla fleshy orange below, cream above. Stigma white. Fruit green when immature, turning orange.

Sidwell K. 416 26/Oct/1995

With: Etuge. Schoenengerger & Takele

Duplicates at:

New dets to M.Cheek at RBG Kew and HNC BP 1601 Yaounde.

Tesseract 4.0.0 | **K000025814** | Score: 30/32

HERB. HORT. KEW.

ROYAL BOTANIC GAR

MINER

A
& PR
Av " He
3 z
. iY
L O 5
) r
p LL g N
{
. \$HrP < '

y LI CLA ¥
J :
i"
()
>
Q |
G8
(0)
a.
=
Oo) |
= |
p_
-
- 2
(&)
THE PLANTS OF WESTERN
CAMEROON
Recorded on Database (K) 1992-
Psychofifa geophylax Cheek & Sonké
(Psych. sp. B aff gabonica)
Cited in protologue
DET Cheek
Jan 2007
cesssesasesanse
FLORA OF WESTERN CAMEROON
RBG Kew & Herbar National du Cameroun
support by Darwin Initiative & Earthwatch
Rubiaceae

Psychotria

Division: Kupe-Muanenguba South West

Gazette: Nyasoso

LongLat: N; E Alt: 1190m

Above Nyasoso on Max's trail up Mount Kupe. Montane forest with canopy to 35m tall. Many stands of Marantaceae. Volcanic soils.

Large shrub to 2.5m tall. Leaves shiny, nerves depressed above. Stipules green, very broad: flowers orange, clustered in dense creamy coloured heads. Buds orange, only a few flowers open per head. Sepals brown: corolla fleshy orange below, cream above. Stigma white. Fruit green when immature, turning orange.

Sidwell K. 416 26/Oct/1995

With: Etuge, Schoenengerger & Takele

Duplicates at:

New dets to M.Cheek at RBG Kew and HNC BP 1601 Yaounde.

Tesseract 3.0.51 | **K000025814** | Score: 24.5/32

HERB HORT KEW

ROYAL BOTANIC GAR

1111111111111111111'1111115

W.

' 7 '9.

x . ,

'

l»

. 4' . .

' l

'1 4' ' .

1

9 "x . ' - /

l

l F A '4 ' »

{fl '

.0

d)

>

L

(D

m

(D

L

4.:

.C

CD

':

>~.

O.

O

0

THE PLANTS OF WESTERN

CAMEROON

Recorded on Database (K) 1992

Psychofla geophylax **Cheek & Sonke**

(Psych ép B aff gabonica)

Cited in protologue

DET Cheek Jan 2007

H OLOTVPL

1457 61101140 **geophylax Cheek** +§oml<e

DET 20

I

FLORA OF WESTERN CAMEROON

RBG Kew & **Herbler** National du Cameroun

support by Darwm Initiative & Earthwatch

Rubiaceae HOLOTV 1715:

PSVC/zotiza 91> B 4'1 'LLCA

bet C&MCL VYDOL

Division Kupe Muanenguba South West

Gazette Nyasoso

LongLat N E Alt 1190m

Above Nyasoso on Max 3 trail up Mount Kupe Montana

forest with canopy to 35m tall Many stands of

Marantaceae Volcanic soils

Large shrub to 2.5m tall Leaves shiny nerves depressed

above Stipules green very broad flowers orange clustered

m dense creamy coloured heads Buds orange only a few

flowers open per head Sepals brown corolla fleshy orange

below cream above Stigma white Fruit green when

immature turning orange
Sidwell K 416 26/Oct/1995
With htuge Sulloencngugel 6L lakele
Dupliutus u k Lu'ol YA' S(k(NAG! BR; r10. P
EKLU arm 1AA EA C&Jfi
. \u\ dCLS 10 M Check it RBU lxew 1nd HNC BP 1601 \ wounds

ABBY FineReader Engine 12 | **K000025814** | Score: 30/32

HERB. HORT. KEW.
ROYAL BOTANIC GARDENS KEW
<000025814

I
Cheek & Sonke
Jan 2007

Cheek -t-Sonke
OX
DET
20

!
1

Rubiaceae

f.
Division:
South West
Gazette:

26/Oct/1995

i
eV
*

c
■

CO

co

/

co

04

4 .

r

E

o o

0)

f

o

cn

CD

Kupe-Muanenguba

Nyasoso

◆

f

Q5<^

LongLat: N; E Alt: 1] 90m

Above Nyasoso on Max's trail up Mount Kupe. Montane

Microsoft OneNote 2013 | **K000025814** | Score: 31.5/32

HERB. HORT. KEW

ROYAL BOTANIC GARDENS KEW

K000025814

SidLL'e(l

THE PLANTS OF WESTERN
CAMEROON

Recorded on Database (K) 1992-

Ps ch01a eo h lax **Cheek & Sonké**

(Psych. **p.** B aff gabonica)

Cited in protologue

DET Cheek

clnö+riQ geo?

lax

DET.....

Jan 2007

Chee14 +

20

FLORA OF WESTERN CAMEROON

RBG Kew & Herbar National du Cameroun

support by Darwin Initiative & Earthwatch

Rubiaceae

Psychotria s

Kupe-Muanenguba

Division.

Nyasoso

Gazette:

LongLat: N; E

A OCOTY/E

CUCA

South West

Alt: 11 90m

Above Nyasoso on Max's trail up Mount Kupe. Montane forest with canopy to 35m tall. Many stands of Marantaceae. Volcanic soils.

Large shrub to 2.5m tall. Leaves shiny, nerves depressed above. Stipules green, very broad; flowers orange, clustered in dense creamy coloured heads. Buds orange, only a few flowers open per head. Sepals brown; corolla fleshy orange below, cream above. Stigma white. Fruit green when immature, turning orange.

Sidwell K. 416

With: Etuge, Schoenengerger & Takele

26/Oct/1995

Duplicates at: **WAG t**

New to M.Cheek at RBG Kew and HNC BP 1601 Yaounde.

2.2 Line Correctness with Segmented Images

2.2.1 Segments of image “B 10 0002520”

	bcd01 (/2)	lbl01 (/1)	lbl02 (/8)	lbl03 (/11)	lbl04 (/1)	lbl06 (/1)
Tesseract 4.0.0	1	1	4.5	10	0	1
Tesseract 3.0.51	1	1	4.5	7	0	1
ABBYY FineReader Engine 12	2	0	6	9.5	0	1
Microsoft OneNote 2013	1.5	0	5	8	0	1

2.2.1.1 Segment "B 10 0002520_bcd01"



Manual transcription of **B 10 0002520_bcd01** | Number of lines: 2

Mus. Bot. Berol.
B 10 0002520

Tesseract 4.0.0 | **B 10 0002520_bcd01** | Score: 1/2

Mus. Bot. Berol.
HL

Tesseract 3.0.51 | **B 10 0002520_bcd01** | Score: 1/2

Mus Bot Berol
Mummmmnlnygigugmgwmmmlwm

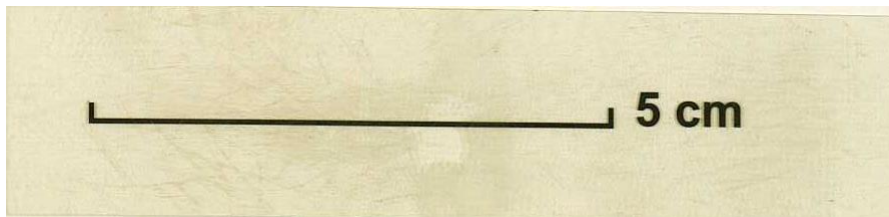
ABBYY FineReader Engine 12 | **B 10 0002520_bcd01** | Score: 2/2

Mus. Bot. Berol.
B 10 0002520

Microsoft OneNote 2013 | **B 10 0002520_bcd01** | Score: 1.5/2

Muş. Bot. Berol.
B 10 0002520

2.2.1.2 Segment "B 10 0002520_lbl01"



Manual transcription of B 10 0002520_lbl01 | Number of lines: 1

5cm

Tesseract 4.0.0 | B 10 0002520_lbl01 | Score: 1/1

1 5cm

Tesseract 3.0.51 | B 10 0002520_lbl01 | Score: 1/1

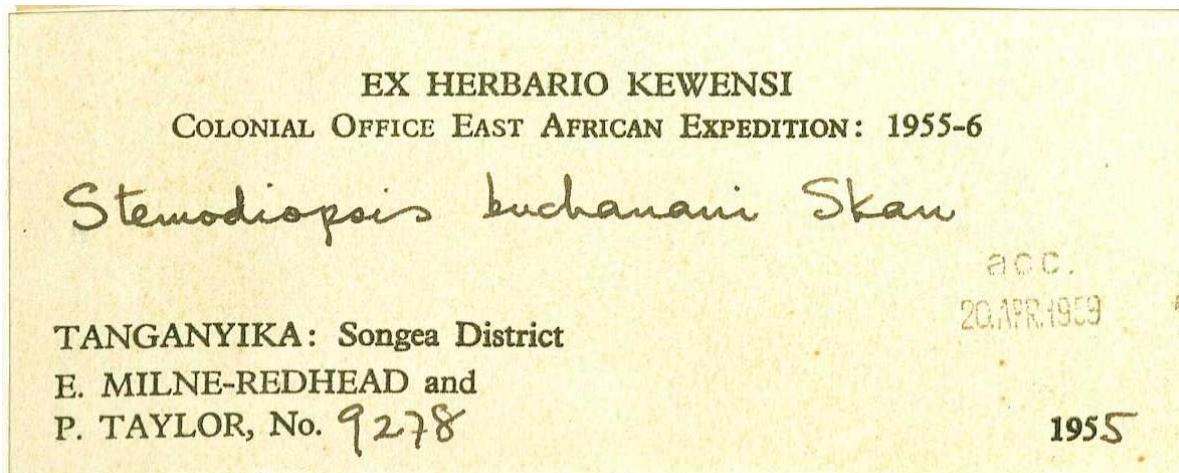
-5cm

ABBYY FineReader Engine 12 | B 10 0002520_lbl01 | Score: 0/1

in
E
o

Microsoft OneNote 2013 | B 10 0002520_lbl01 | Score: 0/1

2.2.1.3 Segment "B 10 0002520_lbl02"



Manual transcription of B 10 0002520_lbl02 | Number of lines: 8

EX HERBARIO KEWENSI
COLONIAL OFFICE EAST AFRICAN EXPEDITION: 1955-6
TANGANYIKA: Songea District
E. MILNE-REDHEAD and
P.TAYLOR, No.
195
acc.
20.APR.1959

Tesseract 4.0.0 | B 10 0002520_lbl02 | Score: 4.5/8

EX HERBARIO KEWENSI
CoLoNIAL **Orrice** EAST AFRICAN EXPEDITION: 1955-6
TANGANYIKA: Songea District
E. MILNE-REDHEAD and
P. TAYLOR, No. 123%

Tesseract 3.0.51 | B 10 0002520_lbl02 | Score: 4.5/8

EX HERBARIO KEWENSI
COLONIAL OFFICE EAST AFRICAN **Expanmon** 1955 6
TANGANYIKA Songea District
E MILNE REDHEAD and
P TAYLOR No fizq'af

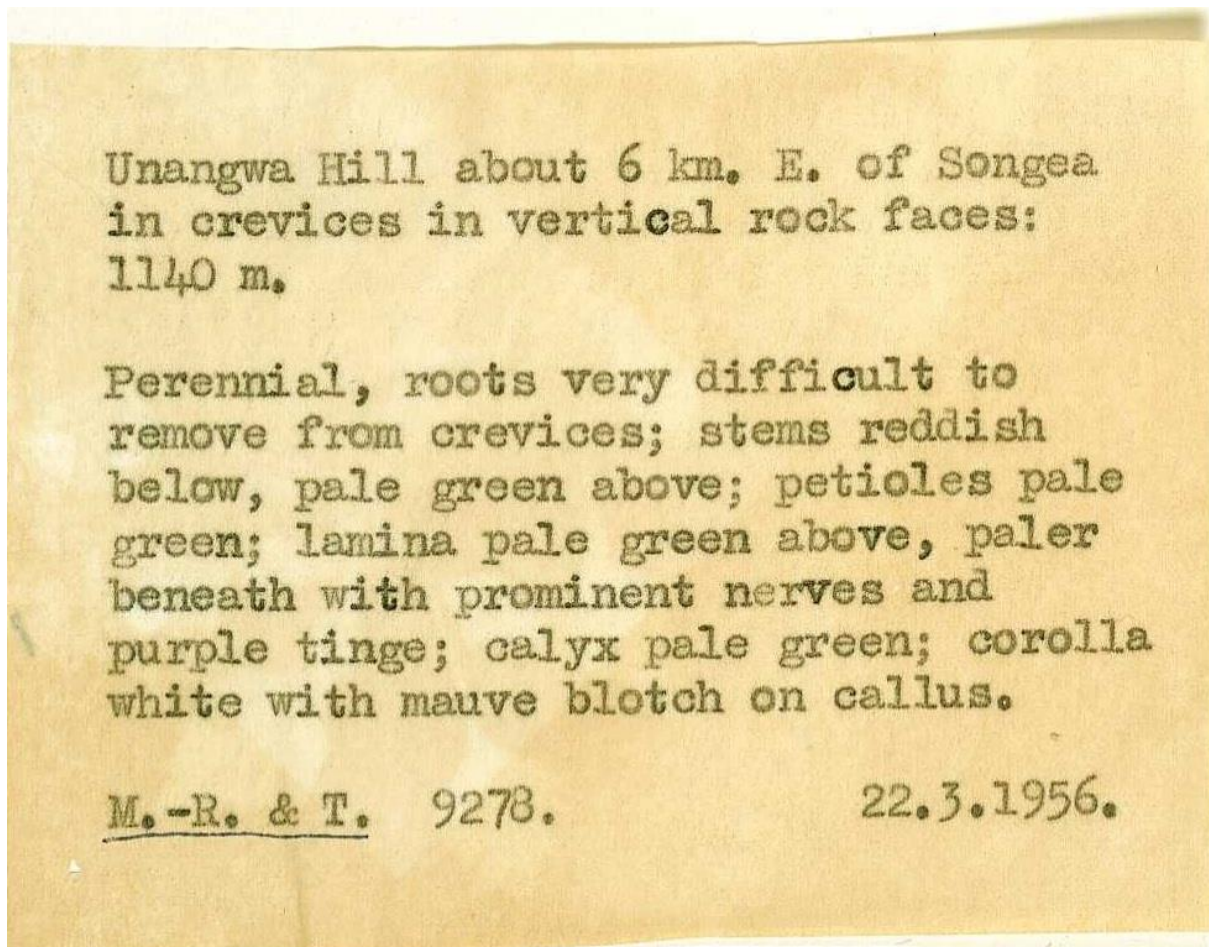
ABBYY FineReader Engine 12 | **B 10 0002520_ibl02** | Score: 6/8

A
1955"
EX HERBARIO KEWENSI
Colonial Office East African Expedition: 1955-6
TANGANYIKA: Songea District
E. MILNE-REDHEAD and
P. TAYLOR, No. ZSfif
a c c.
2G.W9C9

Microsoft OneNote 2013 | **B 10 0002520_ibl02** | Score: 5/8

EX HERBARIO KEWENSI
COLONIAL OFFICE EAST AFRICAN EXPEDITION: 1955-6
TANGANYIKA: Songea District
E. MILNE-REDHEAD and
P. TAYLOR, No.

2.2.1.4 Segment "B 10 0002520_ibl03"



Manual transcription of B 10 0002520_ibl03 | Number of lines: 11

Unangwa Hill about 6 km, E. of Songea
in crevices in vertical rock faces:
1140 m,
Perennial, roots very difficult to
remove from crevices; stems reddish
below, pale green above; petioles pale
green; lamina pale green above, paler
beneath with prominent nerves and :
purple tinge; calyx **pele** green; corolla
white with mauve blotch on callus
Me=Re & T. 9278. 22,3+1956,

Tesseract 4.0.0 | B 10 0002520_ibl03 | Score: 10/11

Unangwa Hill about 6 km, E. of Songea

in crevices in vertical rock faces:
1140 m,
Perennial, roots very difficult to
remove from crevices; stems reddish
below, pale green above; petioles pale
green; lamina pale green above, paler
beneath with prominent nerves and :
purple tinge; calyx **pele** green; corolla
white with mauve blotch on callus
Me=Re & T. 9278. 22,3+1956,

Tesseract 3.0.51 | B 10 0002520_ibl03 | Score: 7/11

Unangwa Hill about 6 km. E of Songea
1n crevices in vertieal rock faces
1140 m
Perennial, roots very difficult to
remove from crevices, stems reddish
below, **pala** green above; patioles pale
green, lamina. pale green above, **yaler**
beneath with prominent **harms** and
purple tinge, calyx pale green, **mmla**
white with mauve blotch **On** callus.
u, R. & T 9278 22 3 1956.

ABBYY FineReader Engine 12 | B 10 0002520_ibl03 | Score: 9.5/11

22.3.1956.
9278.
M.-R. & T.
Perennial, roots very difficult to
remove from crevices; stems reddish
below, pale green above; petioles pale
green; lamina pale green above, paler
beneath with prominent nerves and
purple tinge; calyx pale green; corolla
white with mauve blotch on callus.
Unangwa Hill about 6 **lon**. E. of Songea
in crevices in vertical rock faces:

1140 m.

Microsoft OneNote 2013 | B 10 0002520_ibl03 | Score: 8/11

TJnangwa Eli 11 about 6 lcm. E. of Songea

in crevices in vertical rock faces;

Perenniäl, roots very difficult to

remove from crevices; **sterns** reddish

below, pale green above; petioles pale

green; **laxnina** pale green above, paler

beneath with prominent nerves and

purple tinge; calyx pale green; corolla

white with mauve blotch on callus.

& TO 927B.

22.3.1956.

2.2.1.5 Segment “B 10 0002520_lbl04”



Manual transcription of **B 10 0002520_lbl04** | Number of lines: 1

Image 2001

Tesseract 4.0.0 | **B 10 0002520_lbl04** | Score: 0/1

Tesseract 3.0.51 | **B 10 0002520_lbl04** | Score: 0/1

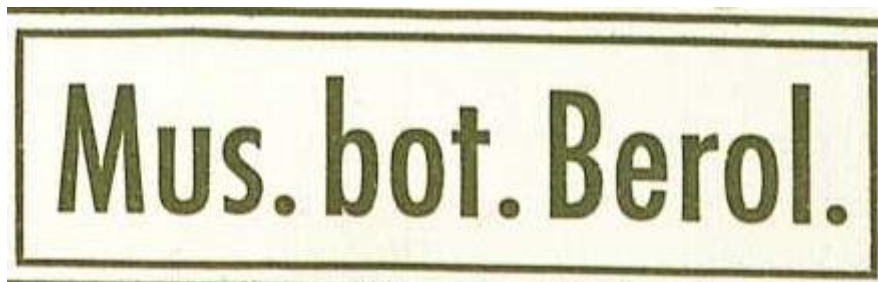
ABBYY FineReader Engine 12 | **B 10 0002520_lbl04** | Score: 0/1

■minimi

Microsoft OneNote 2013 | **B 10 0002520_lbl04** | Score: 0/1

ma e

2.2.1.6 Segment "B 10 0002520_lbl06"



Manual transcription of B 10 0002520_lbl06 | Number of lines: 1

Mus. bot. Berol.

Tesseract 4.0.0 | B 10 0002520_lbl06 | Score: 1/1

Mus. bot. Berol.

Tesseract 3.0.51 | B 10 0002520_lbl06 | Score: 1/1

Mus bot Berol

ABBYY FineReader Engine 12 | B 10 0002520_lbl06 | Score: 1/1

Mus. bot. Berol.

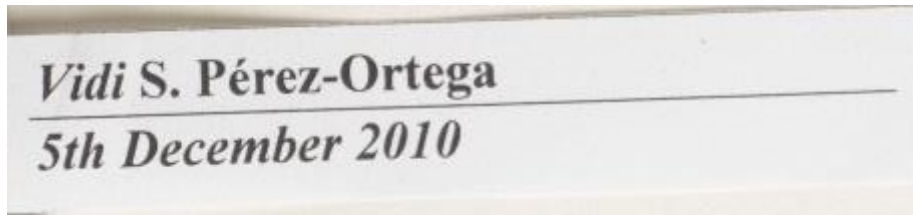
Microsoft OneNote 2013 | B 10 0002520_lbl06 | Score: 1/1

MUS. bot. Berol.

2.3.1 Segments of image “BM000500117”

Image segments [BM000500117]	bcd01 (/3)	lbl03 (/2)	lbl04 (/8)	lbl07 (/1)
Tesseract 4.0.0	1	0	8	1
Tesseract 3.0.51	1	0	5.5	1
ABBYY FineReader Engine 12	3	0	7.5	1
Microsoft OneNote 2013	3	0	7.5	0

2.3.1.2 Segment “BM000500117_lbl03”



Manual transcription of **BM000500117_lbl03** | Number of lines: 2

Vidi S. Pérez-Ortega
5th December 2010

Tesseract 4.0.0 | **BM000500117_lbl03** | Score: 0/2

0102 42quia2ad YI§
en
©30310-23.19d °S PIA
Ce -----

Tesseract 3.0.51 | **BM000500117_lbl03** | Score: 0/2

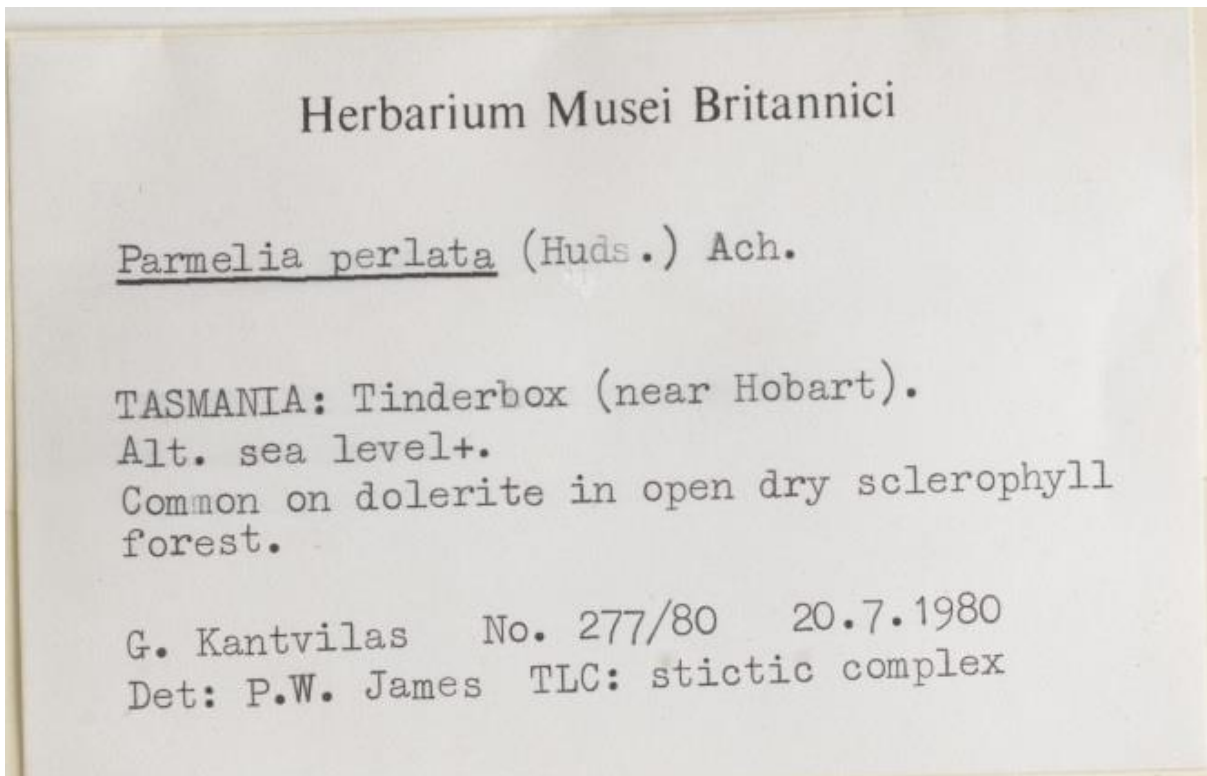
Sam Luefimumfi 5m
\\
awatO N2?— m :2:
III“

ABBYY FineReader Engine 12 | **BM000500117_lbl03** | Score: 0/2

" -CJ
s
Cl
■s
l s
£
l
Oh
l C/5
3
&

Microsoft OneNote 2013 BM000500117_lbl03 Score: 0/2

2.3.1.3 Segment "BM000500117_lbl04"



Manual transcription of **BM000500117_lbl04** | Number of lines: 8

Herbarium Musei Britannici
Parmelia perlata (Huds.) Ach.
TASMANIA: Tinderbox (near Hobart).
Alt. sea level+.
Common on dolerite in open dry sclerophyll
Forest.
G.Kantvilas No. 277/80 20.7.1980
Det: P.W. James TLC: stictic complex

Tesseract 4.0.0 | **BM000500117_lbl04** | Score: 8/8

Herbarium Musei Britannici
Parmelia perlata (Huds.) Ach.
TASMANIA: Tinderbox (near Hobart).
Alt. sea level+.
Common on dolerite in open dry sclerophyll
forest.
G. Kantvilas No. 277/80 20.7.1980
Det: P.W. James TLC: stictic complex

Tesseract 3.0.51 | **BM000500117_ibl04** | Score: 5.5/8

Herbarium Musei Britannici
Parmella Eerlata (Hue) Ash
TASMANIA Tinderbox (near Hobart)
Alt sea level+
Common on dolerite :Ln open dry sclerophyll
forest
G Kantvilas No 277/80 20 7 1980
Det P.W. James TLC stictic complex

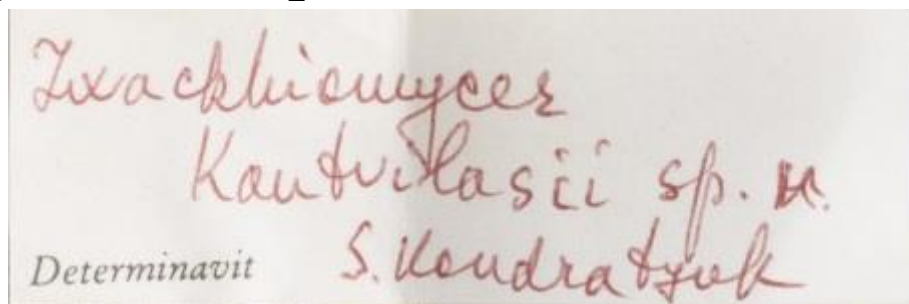
ABBYY FineReader Engine 12 | **BM000500117_ibl04** | Score: 7.5/8

Herbarium Musei Britannici
Parmelia perlata (Hudt..) Ach.
TASMANIA: Tinderbox (near Hobart).
Alt. sea level+.
Common on dolerite in open dry sclerophyll
forest.
G. Kantvilas No. 277/80 20.7-1980
Det: P.W. James TLC: stictic complex

Microsoft OneNote 2013 | **BM000500117_ibl04** | Score: 7.5/8

Herbarium Musei Britannici
Parmelia perlata (Huds) Ach.
TASMANIA: Tinderbox (near Hobart) .
Alt. sea level+.
Common on dolerite in open dry sclerophyll
forest.
g. Kantvilas No. 277/80 20.7. 1980
Det: P.W. James TLC: stictic complex

2.3.1.4 Segment "BM000500117_lbl07"



Manual transcription of **BM000500117_lbl07** | Number of lines: 1

Determinavit

Tesseract 4.0.0 | **BM000500117_lbl07** | Score: 1/1

Determinavit

Tesseract 3.0.51 | **BM000500117_lbl07** | Score: 1/1

Determinavit

ABBYY FineReader Engine 12 | **BM000500117_lbl07** | Score: 1/1

Determinavit

t cuu^£

\$ ll

Microsoft OneNote 2013 | **BM000500117_lbl07** | Score: 0/1

Determinavzt

2.4.1 Segments of image “E00015443”

Image segments [E00015443]	bcd01 (/3)	lbl01 (/1)	lbl03 (/6)
Tesseract 4.0.0	2	1	5
Tesseract 3.0.51	0	1	3.5
ABBYY FineReader Engine 12	2	1	6
Microsoft OneNote 2013	2	1	5

2.4.1.1 Segment "E00015443_bcd01"



Manual transcription of E00015443_bcd01 | Number of lines: 3

ROYAL BOTANIC GARDEN
EDINBURGH
E00015443

Tesseract 4.0.0 | E00015443_bcd01 | Score: 2/3

ROYAL BOTANIC GARDEN
EDINBURGH

Tesseract 3.0.51 | E00015443_bcd01 | Score: 0/3

AAAAAAAAAAAAAAAAAAAA
EEEEEEEEEE

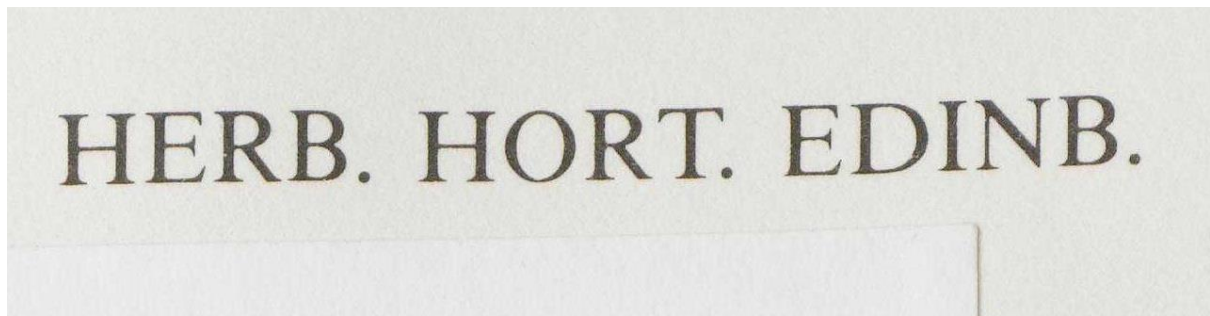
ABBYY FineReader Engine 12 | E00015443_bcd01 | Score: 2/3

E00015443
ROYAL BOTANIC GARDEN
EDINBURGH

Microsoft OneNote 2013 | E00015443_bcd01 | Score: 2/3

E00015443
ROYAL BOTANIC GARDEN
EDINBURGH

2.4.1.2 Segment "E00015443_lbl01"



Manual transcription of E00015443_lbl01 | Number of lines: 1

HERB. HORT. EDINB.

Tesseract 4.0.0 | E00015443_lbl01 | Score: 1/1

HERB. HORT. EDINB

Tesseract 3.0.51 | E00015443_lbl01 | Score: 1/1

HERB HORT EDINB

{

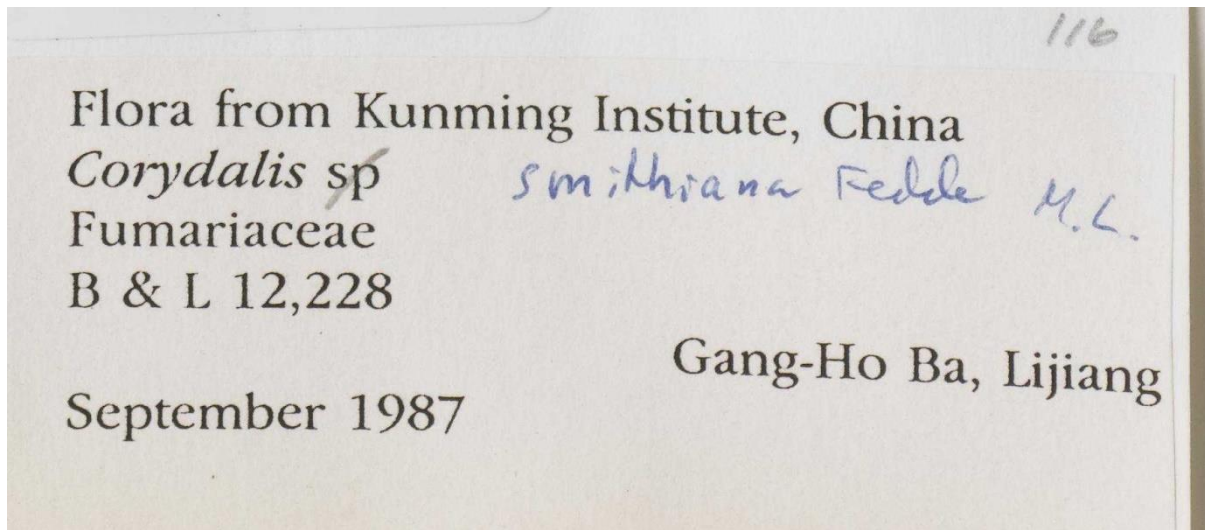
ABBYY FineReader Engine 12 | E00015443_lbl01 | Score: 1/1

HERB. HORT. EDINB.

Microsoft OneNote 2013 | E00015443_lbl01 | Score: 1/1

HERB. HORT. EDINB

2.4.1.3 Segment "E00015443_lbl03"



Manual transcription of **E00015443_lbl03** | Number of lines: 6

Flora from Kunming Institute, China
Corydalis
Fumariaceae
B & L 12,228
Gang-Ho Ba, Lijiang
September 1987

Tesseract 4.0.0 | **E00015443_lbl03** | Score: 5/6

Flora from Kunming Institute, China
Corydalis sp smithrana Rede 4,
Fumariaceae we,
B&1L12228
Gang-Ho Ba, Lijiang
September 1987

Tesseract 3.0.51 | **E00015443_lbl03** | Score: 3.5/6

Flora from Kunming Institute, China
COlj'dallS Spy J m IU'h {4 H» Rd; 1,;
Fumaridceae (L
B & L 12 228
Gang HO Ba, Lijiang
September 1987

ABBY FineReader Engine 12 | **E00015443_ibl03** | Score: 6/6

Ara ReAIL
Gang-Ho Ba, Lijiang
September 1987
Flora from Kunming Institute, China
Corydalis sp"
Fumariaceae
B & L 12,228

Microsoft OneNote 2013 | **E00015443_ibl03** | Score: 5/6

Flora from Kunming Institute, China
Col)'dalis sf S m
Fumariaceae
B & L 12,228
September 1987
Gang-Ho Ba,
Lijiang

2.5.1 Segments of image “EIG.2770”

Image segments [EIG.2770]	bcd01 (/4)	lbl02 (/7)	lbl03 (/4)
Tesseract 4.0.0	3	7	3
Tesseract 3.0.51	2	6	2.5
ABBYY FineReader Engine 12	3	6.5	3
Microsoft OneNote 2013	2.5	6.5	2.5

2.5.1.1 Segment "EIG.2770_bcd01"



Manual transcription of **EIG.2770_bcd01** | Number of lines: 4

Digitarium
http://id.luomus.fi/
EIG.2770
2013-04-02

Tesseract 4.0.0 | **EIG.2770_bcd01** | Score: 3/4

http://id.luomus.fi/ |
EIG.2770
2013-04-02
Fi
o
£
3
To
ji]
--
=
O

Tesseract 3.0.51 | **EIG.2770_bcd01** | Score: 2/4

http //id luomus fi/
EIG 2770
2013 04 02
533
E!
E

g
L-
1'5
9.-
g:
o

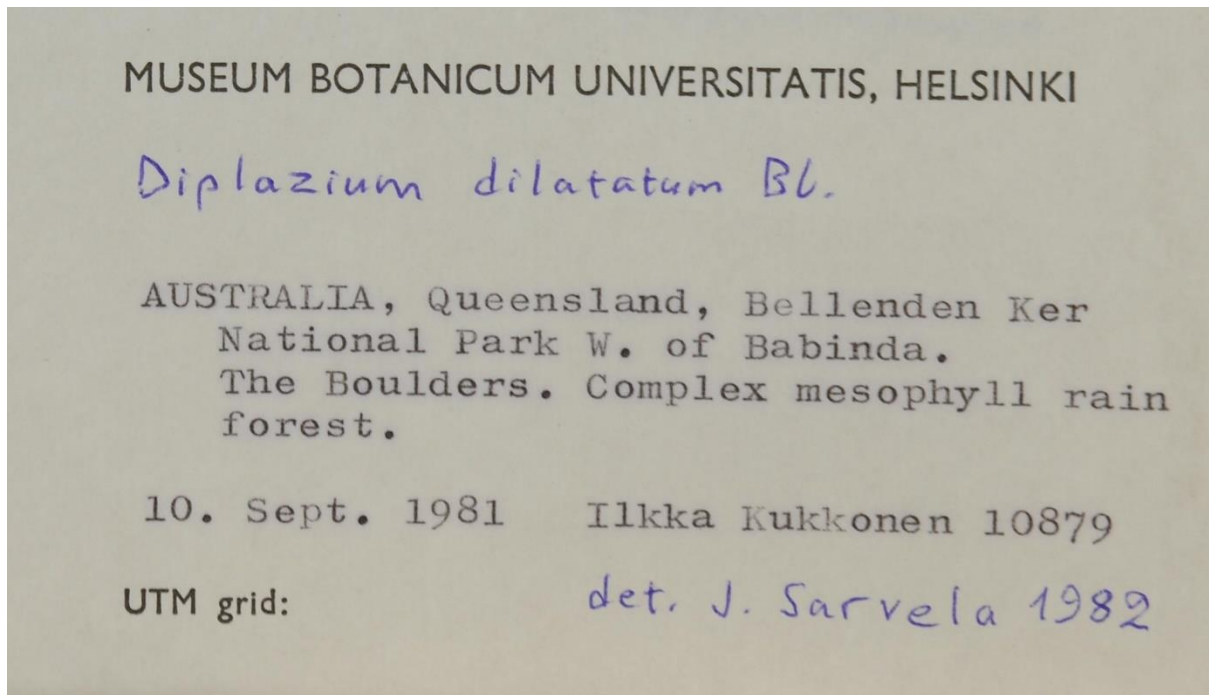
ABBYY FineReader Engine 12 | **EIG.2770_bcd01** | Score: 3/4

<http://id.luomus.fi/>
EIG.2770
2013-04-02

Microsoft OneNote 2013 | **EIG.2770_bcd01** | Score: 2.5/4

<http://id.luomus.fi/>
EIG.2770
2013-04-02

2.5.1.2 Segment "EIG.2770_ibl02"



Manual transcription of **EIG.2770_ibl02** | Number of lines: 7

MUSEUM BOTANICUM UNIVERSITATIS, HELSINKI
AUSTRALIA, Queensland, Bellenden Ker
National Park W. of Babinda.
The Boulders. Complex mesophyll rain
Forest.
10. Sept. 1981 Ilkka Kukkonen 10879
UMT grid:

Tesseract 4.0.0 | **EIG.2770_ibl02** | Score: 7/7

MUSEUM BOTANICUM UNIVERSITATIS, HELSINKI
AUSTRALIA, Queensland, Bellenden Ker
National Park W. of Babinda.
The Boulders. Complex mesophyll rain
forest,
10. Sept. 1981 Ilkka Kukkonen 10879
UTM grid: det

Tesseract 3.0.51 | **EIG.2770_ibl02** | Score: 6/7

MUSEUM BOTANICUM UNIVERSITATIS HELSINKI

Dip [02(um <11 latafurn BL
AUSTRALIA, Queensland, Bcllenden ker
National Park W. of Babinda.
The Boulders Complex mesophyll rain
forest.
10 bept 1981 Illka Lullonen 10879
UTM grid clef J SuFVQIQ {}82

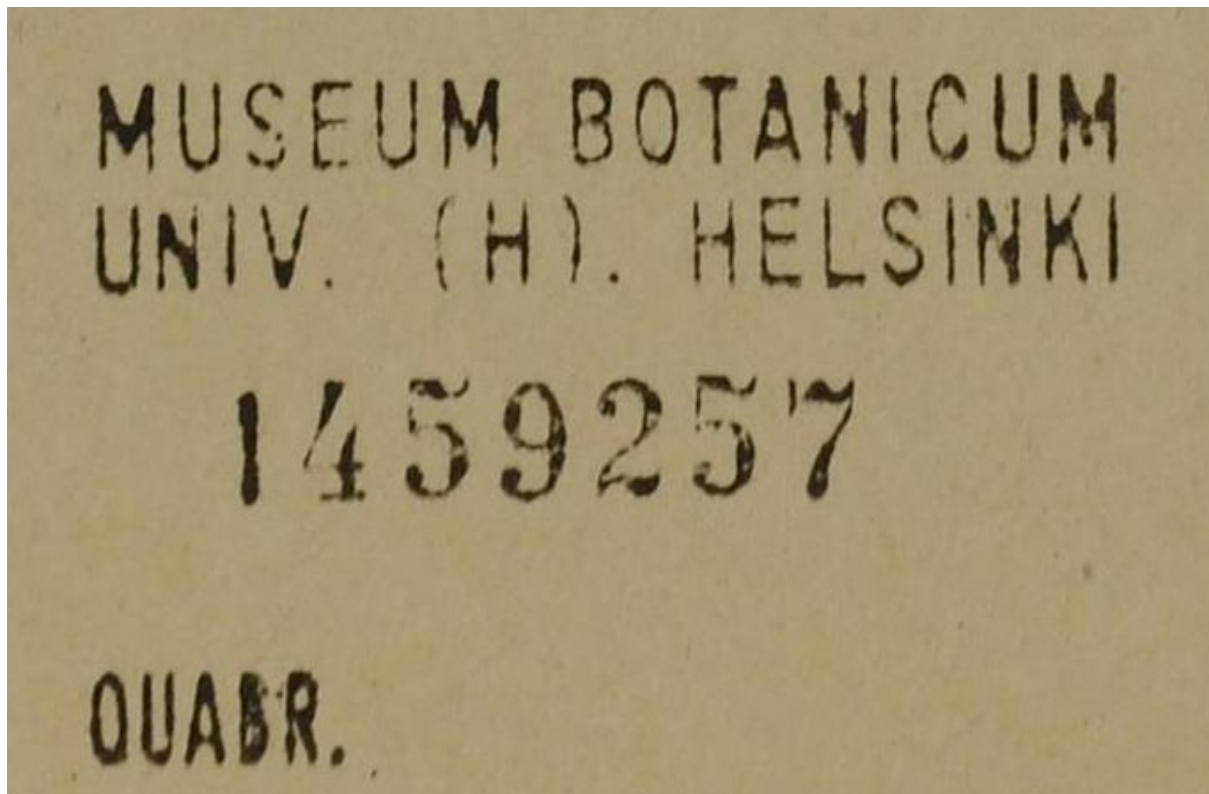
ABBYY FineReader Engine 12 | EIG.2770_ibl02 | Score: 6.5/7

MUSEUM BOTANICUM UNIVERSITATIS, HELSINKI
J i i C K L,
C* I p i a Zi U 4<V\
10. Sept. 1981
Ilkka Kukkonen IO879
J. WvgJa 4382
UTM grid:
AUSTRALIA, Queensland, Bellenden Ker
National Park W. of Babinda.
The Boulders. Complex mesophyll rain
forest.

Microsoft OneNote 2013 | EIG.2770_ibl02 | Score: 6.5/7

MUSEUM BOTANICUM UNIVERSITATIS, HELSINKI
AUSTRALIA, Queensland, Bellenden Ker
National Park W. of Babinda
The Boulders
forest.
10. Septe 1981
UTM grid:
Complex mesophy11 rain
Ilkka Ikukkonen 10879
Jet, J.

2.5.1.3 Segment "EIG.2770_lbl03"



Manual transcription of EIG.2770_lbl03 | Number of lines: 4

MUSEUM BOTANICUM
UNIV. (H). HELSINKI
1459257
QUABR.

Tesseract 4.0.0 | EIG.2770_lbl03 | Score: 3/4

MUSEUM BOTANICUM
UNIV. {H}. HELSINKI
}438257
QUABR.

Tesseract 3.0.51 | EIG.2770_lbl03 | Score: 2.5/4

MUSEUM BOTANICUM
UNIV 'H) iiELSIHKI
l4 39237
QUABR

ABBYY FineReader Engine 12 | **EIG.2770_ibl03** | Score: 3/4

OUAtt.
MUSEUM BOTANICUM
UNIV. (H). HELSINKI
1459257

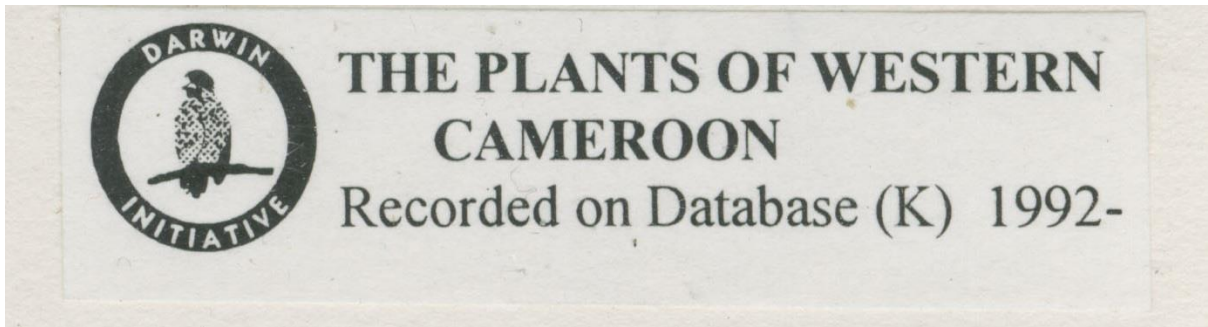
Microsoft OneNote 2013 | **EIG.2770_ibl03** | Score: 2.5/4

MUSEUM BOTANICUM
HELSINKI
UNIV. (H).
OUAbR.

2.6.1 Segments of image “K000025814”

Image segments [K000025814]	K000025814_bcd01 (/2)	K000025814_lbl01 (/3)	K000025814_lbl03 (/4)	K000025814_lbl04 (/1)	K000025814_lbl06 (/1)	K000025814_lbl07 (/21)
Tesseract 4.0.0	0.5	3	4	1	1	21
Tesseract 3.0.51	0.5	3	2.5	1	1	18
ABBYY FineReader Engine 12	2	3	3.5	0.5	1	20.5
Microsoft OneNote 2013	2	3	3	0	1	20

2.6.1.2 Segment "K000025814_lbl01"



Manual transcription of **K000025814_lbl01** | Number of lines: 3

THE PLANTS OF WESTERN
CAMEROON
Recorded on Database (K) 1992-

Tesseract 4.0.0 | **K000025814_lbl01** | Score: 3/3

THE PLANTS OF WESTERN
CAMEROON
Recorded on Database (K) 1992-

Tesseract 3.0.51 | **K000025814_lbl01** | Score: 3/3

THE PLANTS OF WESTERN
CAMEROON
Recorded on Database (K) 1992

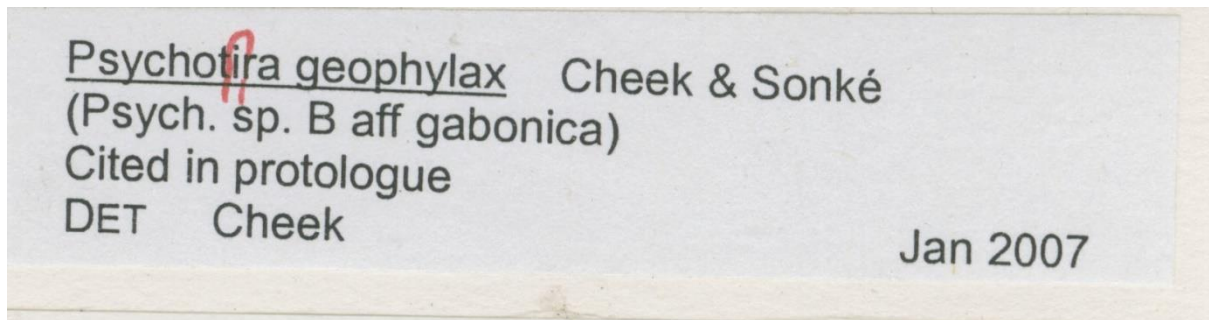
ABBYY FineReader Engine 12 | **K000025814_lbl01** | Score: 3/3

THE PLANTS OF WESTERN
CAMEROON
Recorded on Database (K) 1992-

Microsoft OneNote 2013 | **K000025814_lbl01** | Score: 3/3

* THE PLANTS OF WESTERN
CAMEROON
Recorded on Database (K) 1992-

2.6.1.3 Segment "K000025814_lbl03"



Manual transcription of **K000025814_lbl03** | Number of lines: 4

Psychotira geophylax Cheek & Sonké
(Psych. Sp. B aff gabonica)
Cited in protologue
DET Cheek Jan 2007

Tesseract 4.0.0 | **K000025814_lbl03** | Score: 4/4

Psychofita geophylax Cheek & Sonké
(Psych. 'sp. B aff gabonica)
Cited in protologue
DET Cheek Jan 2007

Tesseract 3.0.51 | **K000025814_lbl03** | Score: 2.5/4

Psvchofla qeophvlax Cheek & Sonke
(Psych 'ép B aff gabonica)
Cited in protologue
DET Cheek Jan 2007

ABBYY FineReader Engine 12 | **K000025814_lbl03** | Score: 3.5/4

4
Jan 2007
Psycho|ira qeophylax Cheek & Sonke
(Psych, sp. B aff gabonica)
Cited in protologue
DET Cheek

Microsoft OneNote 2013 | **K000025814_ibl03** | Score: 3/4

Ps chota eo h lax **Cheek & Sonké**

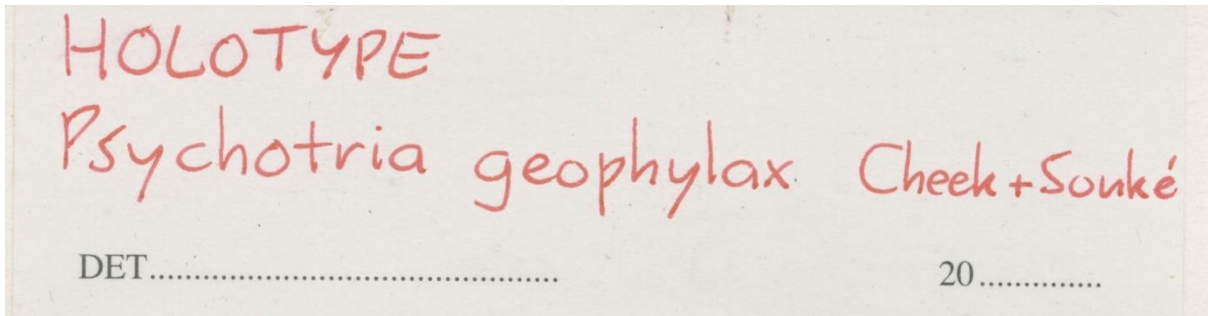
(Psych. p. B aff gabonica)

Cited in protologue

DET Cheek

Jan 2007

2.6.1.4 Segment "K000025814_lbl04"



Manual transcription of K000025814_lbl04 | Number of lines: 1

Det... 20...

Tesseract 4.0.0 | K000025814_lbl04 | Score: 1/1

DET ici sinister 20

Tesseract 3.0.51 | K000025814_lbl04 | Score: 1/1

DET ... 20 ...

ABBYY FineReader Engine 12 | K000025814_lbl04 | Score: 0.5/1

DET...

Microsoft OneNote 2013 | K000025814_lbl04 | Score: 0/1

... ⊥ ∃ CI

⊥ 010H

2.6.1.5 Segment "K000025814_lbl06"



Manual transcription of **K000025814_lbl06** | Number of lines: 1

HERB. HORT. KEW.

Tesseract 4.0.0 | **K000025814_lbl06** | Score: 1/1

HERB. HORT. KEW;

Tesseract 3.0.51 | **K000025814_lbl06** | Score: 1/1

HERB HORT KEW

ABBYY FineReader Engine 12 | **K000025814_lbl06** | Score: 1/1

HERB. HORT. KEW.

Microsoft OneNote 2013 | **K000025814_lbl06** | Score: 1/1

HERB. HORT. KEW.

2.6.1.6 Segment "K000025814_lbl07"

FLORA OF WESTERN CAMEROON
RBG Kew & Herbar National du Cameroun
support by Darwin Initiative & Earthwatch

Rubiaceae

HOLOTYPE

Psychotria sp. B. aff. *gabonica*

Det. Cheek (K) v. 2002

Division: Kupe-Muanenguba

South West

Gazette: Nyasoso

LongLat: N; E

Alt: 1190m

Above Nyasoso on Max's trail up Mount Kupe. Montane forest with canopy to 35m tall. Many stands of Marantaceae. Volcanic soils.

Large shrub to 2.5m tall. Leaves shiny, nerves depressed above. Stipules green, very broad; flowers orange, clustered in dense creamy coloured heads. Buds orange, only a few flowers open per head. Sepals brown; corolla fleshy orange below, cream above. Stigma white. Fruit green when immature, turning orange.

Sidwell K. 416

26/Oct/1995

With: Etuge, Schoenengerger & Takele

Duplicates at: K holo, YA! SCA! WAG, BR, MO, P

BRLV extra YA! EA, CANB

New dets to M.Cheek at RBG Kew and HNC BP 1601 Yaounde.

Manual transcription of **K000025814_ibl07** | Number of lines: 21

FLORA OF WESTERN CAMEROON

RBG Kew & Herbarium National du Cameroun
support by Darwin Initiative & Earthwatch

Rubiaceae

Psychotria

Division: Kupe-Muanenguba South West

Gazette: Nyasoso

LongLat: N; E Alt: 1190m

Above Nyasoso on Max's trail up Mount Kupe. Montane forest with canopy to 35cm tall. Many stands of Marantaceae. Volcanic soils.

Large shrub to 2.5m tall. Leaves shiny, nerves depressed above. Stipules green, very broad; flowers orange, clustered in dense creamy coloured heads. Buds orange, only a few flowers open per head. Sepals brown; corolla fleshy orange below, cream above. Stigma white. Fruit green when immature, turning orange.

Sidwell K. 416 26/Oct/1995

With: Etuge. Schoenengerger & Takele

Duplicates at:

New det. to M. Cheek at RBG Kew and HNC BP 1601 Yaounde.

Tesseract 4.0.0 | **K000025814_ibl07** | Score: 21/21

FLORA OF WESTERN CAMEROON

RBG Kew & Herbarium National du Cameroun
support by Darwin Initiative & Earthwatch

Rubiaceae HOLOT wre

Psychotria Sp. BR df' naa

Det. Clack (2) V.2002

Division: Kupe-Muanenguba South West

Gazette: Nyasoso

LongLat: N; E Alt: 1190m

Above Nyasoso on Max's trail up Mount Kupe. Montane forest with canopy to 35m tall. Many stands of Marantaceae. Volcanic soils.

Large shrub to 2.5m tall. Leaves shiny, nerves depressed above. Stipules green, very broad: flowers orange, clustered

in dense creamy coloured heads. Buds orange, only a few flowers open per head. Sepals brown: corolla fleshy orange below, cream above. Stigma white. Fruit green when immature, turning orange.

Sidwell K. 416 26/Oct/1995

With: Etuge, Schoenengerger & Takele

Duplicates at: "J halo, Ya! Seal Wh | Br, no, FP
oe BRLV Sua ya! EA, Cad

New dets to M.Cheek at RBG Kew and HNC BP 1601 Yaounde.

Tesseract 3.0.51 | **K000025814_ibl07** | Score: 18/21

FLORA OF WESTERN CAMEROON

RBG Kew & **Herbler Natlonal** du Cameroun
support by **Darwm Inmatlve** & Earthwatch

Rubiaceae H O LOT '7 176:

Psychotria g? B 0%, 'L; CA

M GAQQILCL V1001

Division Kupe Muanenguba South West

Gazette Nyasoso

LongLat N E Alt 1190m

Above Nyasoso on Max 3 trail up Mount Kupe Montane
forest with canopy to 35m tall Many stands of
Marantaceae Volcanic soils

Large shrub to 2 5m tall Leaves shiny nerves depressed
above Stipules green very broad flowers orange Clustered
in dense creamy coloured heads Buds orange only a few
flowers open per head Sepals brown corolla fleshy orange
below cream above Stigma white Fruit green when
immature, turning orange

Sidwell K 416 26/Oct/1995

With Etuge **Schoenengerger** & Takele

Dupliuatesat k kn'ol Yh' S(OJ NA&' BK; r10. P

B&Lv Pawn 1M EA C&«J'S

New dets to M Cheek at RBG Kew and HNC BP 1601 Yaounde

ABBY FineReader Engine 12 | **K000025814_ibl07** | Score: 20.5/21

Uocot we

Rubiaceae

Psychotria Sr>. f>.

r

W. CUChCE) V.7DO2-

Division:

Kupe-Muanenguba

South West

Gazette:

Nyasoso

Alt: 1190m

26/Oct/1995

i

Max co.

Large shrub to 2.5m tall. Leaves shiny, nerves depressed above. Stipules green, very broad; flowers orange, clustered in dense creamy coloured heads. Buds orange, only a few flowers open per head. Sepals brown; corolla fleshy orange below, cream above. Stigma white. Fruit green when immature, turning orange.

Sidwell K. 416

With: Etuge, Schoenengerger & Takele

LongLat: N; E

Above Nyasoso on Max's trail up Mount Kupe. Montane forest with canopy to 35m tall. Many stands of Marantaceae. Volcanic soils.

FLORA OF WESTERN CAMEROON

RBG Kew & Herbar National du Cameroun

support by Darwin Initiative & Earthwatch

Duplicates at:

New detts to M.Cheek at RBG Kew and HNC BP 1601 Yaounde.

Microsoft OneNote 2013 | **K000025814_ibl07** | Score: 20/21

FLORA OF WESTERN CAMEROON

RBG Kew & Herbar National du Cameroun

support by Darwin Initiative & Earthwatch

Rubiaceae

Psychotria s. B

Cuecc

Division:

Kupe-Muanenguba

Gazette:

Nyasoso

LongLat: N; E

HO Co TYIIE

South West

Alt: 11 90m

Above Nyasoso on Max's trail up Mount Kupe. Montane forest with canopy to 35m tall. Many stands of Marantaceae. Volcanic soils.

Large shrub to 2.5m tall. Leaves shiny, nerves depressed above. Stipules green, very broad; flowers orange, clustered in dense creamy coloured heads. Buds orange, only a few flowers open per head. Sepals brown; corolla fleshy orange below, cream above. Stigma white. Fruit green when immature, turning orange.

Sidwell K. 416

With: Etuge, Schoenengerger & Takele

26/Oct/1995

Duplicates at: k. (o (SC A (k'A& MO

New **deus** to M.Cheek at RBG Kew and HNC BP 1601 Yaounde.

3 Named Entity Recognition Analysis

3.1 Named Entity Gold Standards

3.1.1 Image “B 10 0002520”

LOCATION	PERSON
Songea District Tanganyika Unangwa Hill Songea	E. Milne-Redhead P. Taylor

Image segment “B 10 0002520_lbl02”

LOCATION	PERSON
Tanganyika Songea District	E. Milne-Redhead P. Taylor

Image segment “B 10 0002520_lbl03”

LOCATION	PERSON
Songea Unangwa Hill	

3.1.2 Image “BM000500117”

LOCATION	PERSON
Tasmania Hobart	E. Kantvilas P.W. Jones

Image segment “BM000500117_lbl04”

LOCATION	PERSON
Tasmania Hobart	E. Kantvilas P.W. Jones

3.1.3 Image “E00015443”

LOCATION	PERSON
China Lijiang	

Image segment “E00015443_lbl03”

LOCATION	PERSON
China Lijiang	

3.1.4 Image “EIG.2770”

LOCATION	PERSON
Australia Helsinki (x2) Babinda Bellenden Ker Queensland	Ilkka Kukkonen

Image segment “EIG.2770_lbl02”

LOCATION	PERSON
Australia Helsinki Babinda Bellenden Ker Queensland	Ilkka Kukkonen

Image segment “EIG.2770_lbl03”

LOCATION	PERSON
Helsinki	

3.1.5 Image “K000025814”

LOCATION	PERSON
Kew Western Cameroon (x2) Cameroun Mount Kype Yaoundé Kupe-Muanenguba Nyasoso (x2)	Cheek (x2) Sonké Etuge Schoenengerger Takele M.Cheek Max Sidwell

Image segment “K000025814_ibl01”

LOCATION	PERSON
Western Cameroon	

Image segment “K000025814_ibl03”

LOCATION	PERSON
	Cheek (x2) Sonké

Image segment “K000025814_ibl07”

LOCATION	PERSON
Western Cameroon Kew Cameroun Kupe-Muanenguba Nyasoso (x2) Mount Kype Yaoundé	Sidwell Etuge Schoenengerger Takele M.Cheek Max

3.2 Stanford NER using OCR text from Whole Images

3.2.1 Image “B 10 0002520”

LOCATION	Lauila
LOCATION	Songea
LOCATION	Songea District
LOCATION	Unangwa Hill
PERSON	Poth

LOCATION		
TP	FP	FN
3	1	1

PERSON		
TP	FP	FN
0	1	2

3.2.2 Image “BM000500117”

LOCATION	Hobart
LOCATION	LONDON
PERSON	P.W. James TLC

LOCATION		
TP	FP	FN
2	0	1

PERSON		
TP	FP	FN
1	0	1

3.2.3 Image “E00015443”

LOCATION	B&L
LOCATION	China
LOCATION	Flora
PERSON	HERB

LOCATION		
TP	FP	FN
1	2	1

PERSON		
TP	FP	FN
0	1	0

3.2.4 Image “EIG.2770”

LOCATION	AUSTRALIA
LOCATION	Babinda
LOCATION	Bellenden Ker MUSEUM BOTANICUM National Park
LOCATION	HELSINKI
LOCATION	HELSINKI
LOCATION	Queensland
LOCATION	UNIVERSITATIS
PERSON	Ilkka Kukkonen

LOCATION		
TP	FP	FN
6	1	0

PERSON		
TP	FP	FN
1	0	0

3.2.5 Image “K000025814”

LOCATION	CAMEROON
LOCATION	CAMEROON
LOCATION	Cameroun
LOCATION	Mount Kupe
LOCATION	Yaounde
PERSON	Cleo
PERSON	HERB
PERSON	Max
PERSON	Montane

LOCATION		
TP	FP	FN
5	0	4

PERSON		
TP	FP	FN
1	3	8

3.3 Stanford NER using OCR text from Segmented Images

3.3.1 Segments of Image “B 10 0002520”

B 10 0002520_lbl02.jpg	LOCATION	Skaw TANGANYIKA
B 10 0002520_lbl02.jpg	LOCATION	Songea District
B 10 0002520_lbl03.jpg	LOCATION	Songea
B 10 0002520_lbl03.jpg	LOCATION	Unangwa Hill

LOCATION		
TP	FP	FN
3	1	0

PERSON		
TP	FP	FN
0	0	2

3.3.2 Segments of Image “BM000500117”

BM000500117_bcd01.jpg	LOCATION	LONDON
BM000500117_lbl04.jpg	LOCATION	Hobart
BM000500117_lbl04.jpg	LOCATION	TASMANIA
BM000500117_lbl04.jpg	PERSON	P.W. James TLC

LOCATION		
TP	FP	FN
3	0	0

PERSON		
TP	FP	FN
1	0	1

3.3.3 Segments of Image “E00015443”

E00015443_bcd01.jpg	LOCATION	ROYAL BOTANIC GARDEN EDINBURGH
E00015443_lbl01.jpg	PERSON	HERB
E00015443_lbl02.jpg	PERSON	INCE
E00015443_lbl03.jpg	LOCATION	Lijiang
E00015443_lbl03.jpg	PERSON	Flora

LOCATION		
TP	FP	FN
1	1	1

PERSON		
TP	FP	FN
0	0	0

3.3.4 Segments of Image “EIG.2770”

EIG.2770_lbl02.jpg	LOCATION	AUSTRALIA
EIG.2770_lbl02.jpg	LOCATION	Babinda
EIG.2770_lbl02.jpg	LOCATION	Bellenden Ker National Park
EIG.2770_lbl02.jpg	LOCATION	HELSINKI
EIG.2770_lbl02.jpg	LOCATION	Queensland
EIG.2770_lbl02.jpg	LOCATION	UNIVERSITATIS
EIG.2770_lbl02.jpg	PERSON	Ilkka Kukkonen
EIG.2770_lbl03.jpg	LOCATION	HELSINKI

LOCATION		
TP	FP	FN
5	1	1

PERSON		
TP	FP	FN
1	0	0

3.3.5 Segments of Image “K000025814”

K000025814_lbl01.jpg	LOCATION	CAMEROON
K000025814_lbl06.jpg	LOCATION	KEW
K000025814_lbl06.jpg	PERSON	HERB
K000025814_lbl07.jpg	LOCATION	Cameroun
K000025814_lbl07.jpg	LOCATION	Mount Kupe
K000025814_lbl07.jpg	LOCATION	Yaounde
K000025814_lbl07.jpg	PERSON	BRLYV Sua
K000025814_lbl07.jpg	PERSON	Clack
K000025814_lbl07.jpg	PERSON	Max
K000025814_lbl07.jpg	PERSON	Montane

LOCATION		
TP	FP	FN
5	0	4

PERSON		
TP	FP	FN
1	4	0

TOTAL

LOCATION		
TP	FP	FN
17	3	6

PERSON		
TP	FP	FN
3	4	3

4 Non-standard Terminology Extraction Analysis

4.1 FlexiTerm Output Using OCR Text From Segmented Images

Coalesced OCR text output from segmented images listed in A3.2. OCR was performed using Tesseract 4.0.0.

Rank	Term representative	Score	Frequency
1	royal botanic gardens	63.1336	60
2	royal botanic	39.6827	61
3	royal botanic garden Edinburgh	35.5238	27
4	MO	31.075	31
5	museum botanicum	20.7944	31
6	horti bot	18.0218	27
7	nationaal herbarium	12.8232	20
8	nationaal herbarium Nederland	11.9016	12
9	mo at bm	9.7041	7
10	natural history	8.7337	14
11	royal botanic gardens ke	8.3178	6
12	herbarium nederland	7.8227	13
13	j. sarvela	7.6246	12
14	center herb	7.278	12
15	imaged gpi	6.9315	11
16	tvaoy royal botanic garden edinburgh	6.4378	4
17	herbier du jardin botanique	6.317	7
18	herbarium bogoriense	6.065	10
19	imaged african plant initiative	5.5452	4
19	nn mo	5.5452	5
19	ethiopian flora project expedition	5.5452	4
19	missouri botanical garden herbarium	5.5452	5
20	york botanical garden	5.4931	6
20	bogor herbarium of new identification	5.4931	5
21	b. bartholomew	4.852	7
21	m. watson	4.852	7
21	m. gilbert	4.852	7
22	flora of china sino-british qinghai expedition	4.8283	3
23	sino-british qinghai expedition	4.3944	7
23	national science foundation	4.3944	4
23	herbier musaum paris	4.3944	5
24	c.r. fraser-jenkins	4.1589	6
24	ilkka kukkonen	4.1589	7
24	royal botanic gardens kew	4.1589	4
24	american museum of natural history	4.1589	4
24	herbier musaum paris p02733867	4.1589	4
25	harvard university	4.0203	7

Appendices (Owen, D. et al. 2020)

26	herbarium vadense	3.4657	5
26	west china	3.4657	5
26	george forrest	3.4657	6
26	plants of Qinghai	3.4657	6
26	paasasai jybuadoo	3.4657	6
26	university of Helsinki	3.4657	5
26	rain forest	3.4657	6
26	utm grid	3.4657	6
26	paris herbier	3.4657	6
26	department of botany	3.4657	6
27	universiteit leiden branch	3.2958	3
27	c. g. pringle	3.2958	3
27	thomas b. croat	3.2958	3
28	paniasal jybuadoo	3.2347	6
29	alpine meadow	2.7726	5
29	arnold arboretum	2.7726	4
29	e.j. strongman	2.7726	4
29	r. steele	2.7726	4
29	voucher specimen	2.7726	5
29	t. Reichstein	2.7726	4
29	flora of Ethiopia	2.7726	8
29	e.j. weeda	2.7726	4
29	a. gray	2.7726	5
29	type specimen	2.7726	5
29	herbier du jardin botanique de eta	2.7726	3
30	paniasal ybuadoo ol	2.1972	3
30	line for computer entry	2.1972	3
31	yushu xian	2.0794	4
31	e side	2.0794	4
31	consolidated scree	2.0794	3
31	institute of botany	2.0794	4
31	academia sinica	2.0794	3
31	r.j.d. mcbeath	2.0794	3
31	flora of taiwan	2.0794	3
31	national taiwan	2.0794	3
31	jaakko sarvela	2.0794	3
31	western australian	2.0794	4
31	guinea expedition	2.0794	3
31	richard archbold	2.0794	3
31	det. Vern	2.0794	3
31	east kalimantan	2.0794	4
31	herbarium wanariset	2.0794	3
31	bukit raya	2.0794	4
31	flora malesiana	2.0794	4
31	east african	2.0794	4
31	garden belgium	2.0794	4

Appendices (Owen, D. et al. 2020)

32	perennial herb	1.7329	4
32	gybuadoo ol	1.7329	4
33	forest shade	1.3863	3
33	ll III	1.3863	3
33	river styx	1.3863	4
33	rock crevices	1.3863	3
33	mixed forest	1.3863	3
33	montane forest	1.3863	3
33	herbario bogoriensi	1.3863	4
33	joint expedition	1.3863	3
33	h.p. nooteboom	1.3863	3
33	herbario kewensi	1.3863	3
33	flora of yemen	1.3863	3
33	r. melville	1.3863	3
34	flora of china	0.6931	4
34	kuo date	0.6931	3
34	g. pringle	0.6931	4