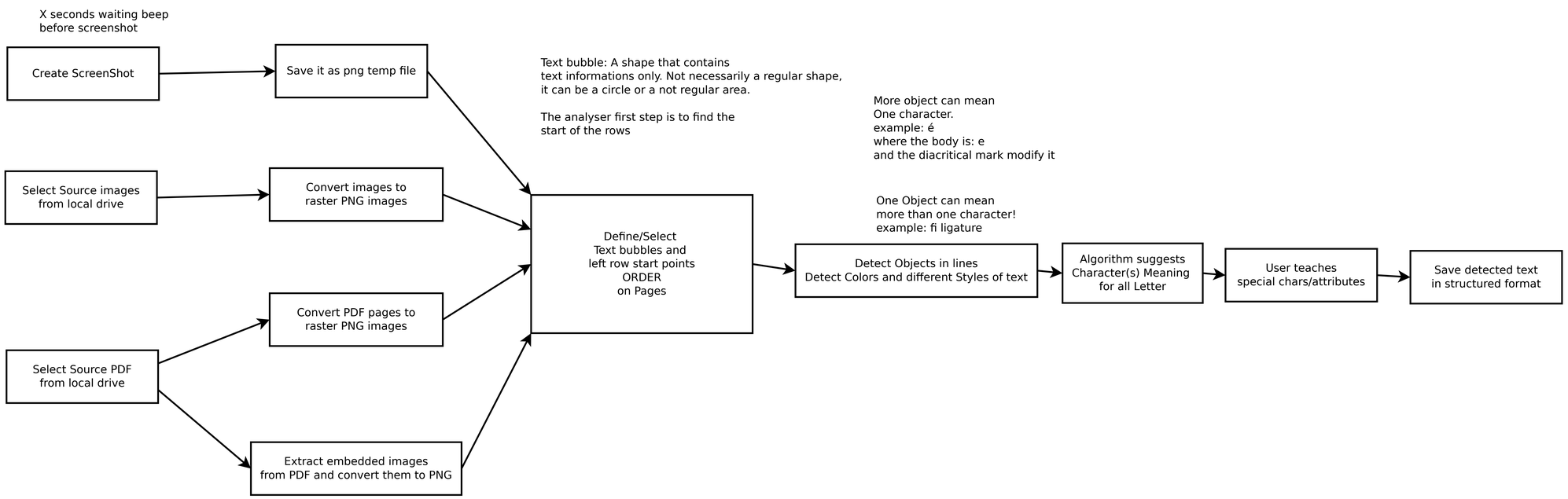
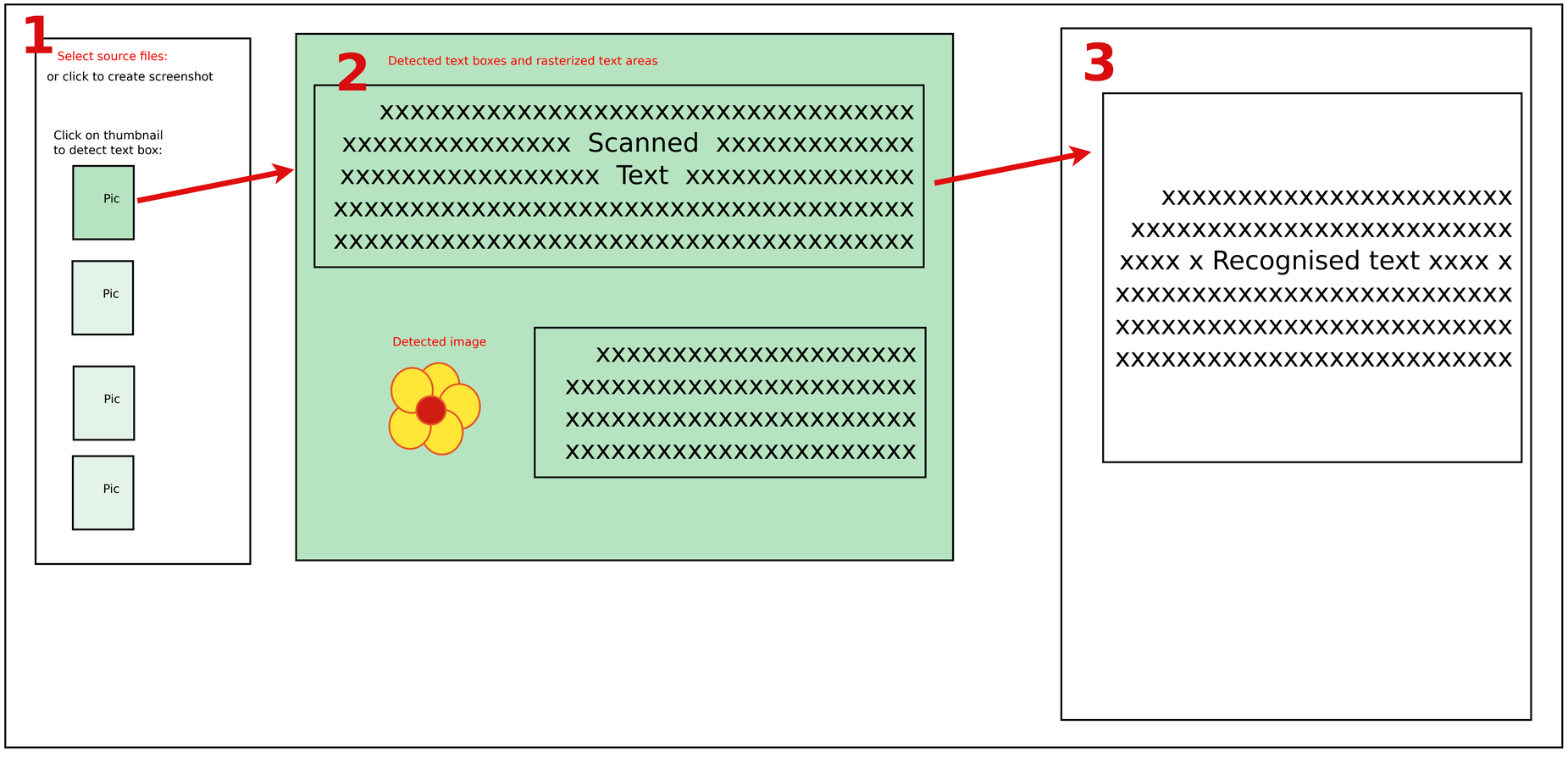

X seconds waiting beep
before screenshot
Create ScreenShot
Save it as png temp file
Select Source images
from local drive
Convert images to
raster PNG images
Convert PDF pages to
raster PNG images
Select Source PDF
from local drive
Extract embedded images
from PDF and convert them to PNG
Text bubble: A shape that contains
text informations only. Not necessarily a regular shape,
it can be a circle or a not regular area.
The analyser first step is to find the
start of the rows
Define/Select
Text bubbles and
left row start points
ORDER
on Pages
More object can mean
One character.
example: é
where the body is: e
and the diacritical mark modify it
One Object can mean
more than one character!
example: fi ligature
Detect Objects in lines
Detect Colors and different Styles of text
Algorithm suggests
Character(s) Meaning
for all Letter
User teaches
special chars/attributes
Save detected text
in structured format


1
Select source files:
or click to create screenshot
Click on thumbnail
to detect text box:
Pic
Pic
Pic
Pic
2
Detected text boxes and rasterized text areas
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxx Scanned xxxxxxxxxxxxx
xxxxxxxxxxxxxxxxx Text xxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Detected image
xxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxx
3
xxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxx
xxxx x Recognised text xxxx x
xxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxx