Document Summary

Filename: Invoice-1D81B628-0001.pdf

Document Type: Invoice

Requested Fields:

- Address
- Phone
- Mobile
- Invoice Number
- Date
- Quantity
- Customer No
- GST
- Total Amount

Extracted Summary:

Date : 03/01/2022
Address: AA Enterprises, 123, main street, Chennai, Tamil Nadu, India - 600009
Phone: 91-44-3261 2121
Mobile: 91-99622 21212
Customer no.: AA/22222/2021
GST No: 1234567890A1ZT

Item	Quantity	Rate	Amount
Minerals	200	3.50	710.00
Paints	300	4.50	1360.00
Hardware	250	10.00	2530.00

Expected Output:
Address: AA Enterprises, 123, main street, Chennai, Tamil Nadu, India - 600009
Phone: 91-44-3261 2121
Mobile: 91-99622 21212
Invoice Number: 1D81B628
Date: 03/01/2022
Quantity: 750
Customer No: AA/22222/2021
GST: 1234567890A1ZT
Total Amount: 4600.00

Actually, I want to extract some fields which are not listed in the output field.

Additional Output:
Shipping Address: AA Enterprises, 123, main street, Chennai, Tamil Nadu, India - 600009
State: Tamil Nadu

Reason:
Shipping Address is the same as the Address field.

State is extracted from the address since it is one of the required fields.

Please note: You need to assume that the fields are not always organized in the document as you see.
Also, your solution should be flexible enough to extract any fields even if not seen in this example.
The document format can change in case to case but the fields are more or less will be the same.
Also if additional fields are required that can be done, but they should be optional.
```

Here's a Python script using regular expressions that can extract the required fields. The script assumes that the

```python
import re

# Document content
document = r"""
Invoice
Invoice number 1D81B628
Date : 03/01/2022
Address: AA Enterprises, 123, main street, Chennai, Tamil Nadu, India - 600009
Phone: 91-44-3261 2121
Mobile: 91-99622 21212
Customer no.: AA/22222/2021
GST No: 1234567890A1ZT
Item                    Quantity    Rate      Amount
Minerals                     200       3.50      710.00
Paints                      300       4.50     1360.00
Hardware                     250       10.00     2530.00
"""

# Define regular expressions for extracting fields
address_re = r'Address:\s*(.*)$'
phone_re = r'Phone:\s*(.*)$'
mobile_re = r'Mobile:\s*(.*)$'
invoice_number_re = r'Invoice number\s*(.*)$'
date_re = r'Date\s*:\s*(.*)$'
quantity_re = r'Quantity\s*(.*)$'
customer_no_re = r'Customer\sno[.:]\s*(.*)$'
gst_number_re = r'GST\sNo:\s*(.*)$'
total_amount_re = r'(?:Item|Total)\s*(?:Amount|:)\s*(.*)'
state_re = r',\s([A-Z]\w*\s*,\s*[A-Z]\w*)'

# Extract fields
address = re.search(address_re, document).group(1)
phone = re.search(phone_re, document).group(1)
mobile = re.search(mobile_re, document).group(1)
```
```

```
invoice_number = re.search(invoice_number_re, document).group(1)
date = re.search(date_re, document).group(1)
quantity = re.search(quantity_re, document).group(1)
customer_no = re.search(customer_no_re, document).group(1)
gst_number = re.search(gst_number_re, document).group(1)
total_amount = re.search(total_amount_re, document).group(1)
state = re.search(state_re, address).group(1)

# Calculate total quantity
quantity = sum(int(x) for x in quantity.split())

# Print output
print('Address:', address)
print('Phone:', phone)
print('Mobile:', mobile)
print('Invoice Number:', invoice_number)
print('Date:', date)
print('Quantity:', quantity)
print('Customer No:', customer_no)
print('GST:', gst_number)
print('State:', state)
print('Total Amount:', total_amount)
print('Shipping Address:', address)
```

This script can be modified to extract additional fields as needed. You can add new regular expressions for each