A Knowledge based Fast PIN code Validation System for Dispatch Sorting of Postal Mail[1]

P.Nagabhushan*, S.A.Angadi** and B.S.Anami***
*Department of Studies in Computer Science, University of Mysore, Mysore
** Department of Studies in Computer Science, University of Mysore, Mysore
and Department of Computer Applications BEC, Bagalkot
*** Department of Computer Science & Engineering, BEC, Bagalkot

Abstract

The postal service that delivers mail to the doorstep of the addressee is to be made efficient by making use of the available technology. Many semi skilled and repetitive sub tasks of postal mail service can be automated to make the service efficient. In this paper a knowledge based computational methodology for destination address validation leading to automatic sorting of mail in a dispatch sorting office is presented. Validation involves checking for consistency among the address components that describe the addressee, with special attention to the PIN code of the distribution post office. The proposed system employs a frame structured knowledge base devised to model the pattern of mail dispatch and the type of addresses encountered. The results of applying the methodology to a sizable set of mail addresses for validation and sorting, at the sorting office of Bagalkot, a district in Karnataka state of South India is presented. The methodology has achieved 90% efficiency in validation and sorting.
Keywords: Automatic Mail Sorting; Knowledge based dispatch sorting; PIN code validation; Frames;

1.0 Introduction

Computer and communication technologies have made many hitherto unthinkable applications a reality. Computer applications are now found in almost all conceivable forms of human endeavor like home, office, business, service industry, manufacturing industry, entertainment industry, education etc. One of the prominent computer applications is the automation of menial tasks. The postal service that delivers written mail and other articles to the doorstep comprises of many menial and repetitive tasks, automation of which makes these tasks efficient. Mail processing/ handling is a very labor intensive task, as the mail is collected from letterbox, sorted according to their destination places and sent to these places in a mailbag. The mail processing task is divided into three major sub tasks as depicted in Figure-1.

```
┌──────────────────┐     ┌──────────────────┐     ┌──────────────────┐
│ Mail Collection  │ ──▶ │   Mail Sorting   │ ──▶ │ Mail Distribution│
└──────────────────┘     └──────────────────┘     └──────────────────┘
```

**Figure-1: Sub tasks of mail processing**

Mail collection involves gathering mail from letter boxes; mail sorting involves classifying and grouping the mail and forwarding it towards the destination. The mail distribution is the delivery of mail to the door step of the addressee.

To make the postal services attractive and efficient there is a need to automate the mail processing/ handling system. The postal automation efforts are reported from United States, Japan, New Zealand, Gulf Countries and China etc. Some of the major companies like National Electric Company (NEC), Siemens, Tritek, Stanford Research International (SRI) and the like provide mail automation solutions. Kanehiro KUBOTA et.al, (1999) have listed equipments used for postal mechanization developed by NEC and used in JAPAN and other countries. The Indian Postal department is also gearing up to use technology for providing efficient postal mail services. The automation of mail collection and distribution involves finding the optimal routes for collection and distribution of mail. Automation of mail sorting requires computerization of its subtasks namely, postal document pre-processing, script identification, optical

character recognition, address transliteration/ translation, address component identification, address validation and address interpretation for dispatch or distribution sorting. The sequence of various sub tasks of mail sorting is brought out in Figure-2.
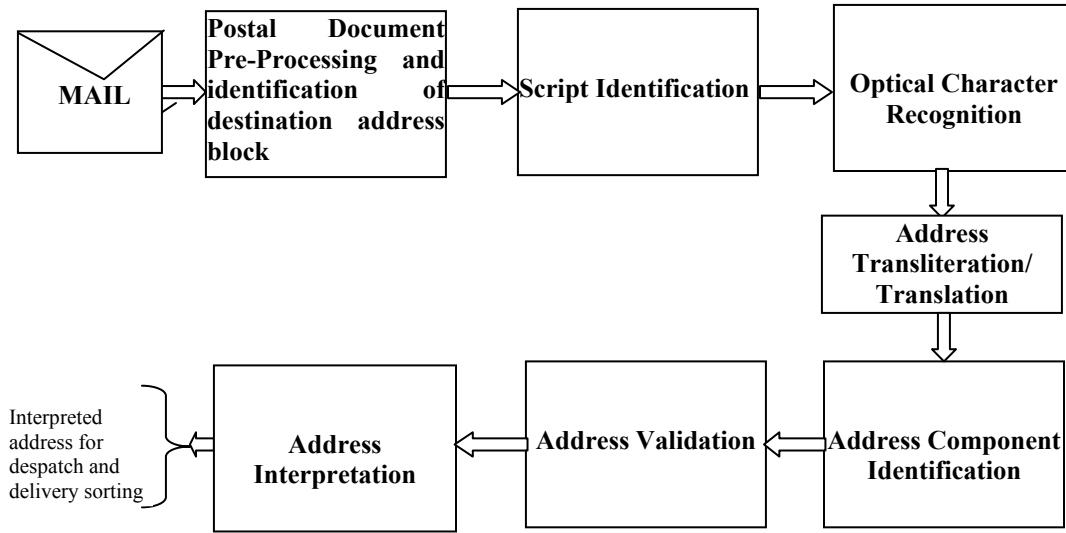
Figure 2: Mail Sorting Sequence

The complete mail sorting task is made up of the sub tasks, which are depicted by separate boxes in Figure-2. The task of postal document pre processing involves image pre processing and identification of destination address block [2]. The next sub task called script identification involves finding the script of the postal destination address (address tokens). The optical character recognition stage recognizes the characters of the destination address. The Address Transliteration/ Translation stage converts the addresses in Indian languages like Kannada, Hindi to English. The Address Component Identification involves extraction of the address components and their identification. The next subtask of mail sorting, which is the subject of this paper, namely address validation involves checking for consistency among address components and their correction, if necessary. The task of address validation assumes more significance in the Indian context as the destination addresses are unstructured and the PIN (Postal Index Number) is not widely known. Also, the postal addresses are written based on a vague geographical knowledge of the destination, which makes address validation an essential component of automatic address interpretation. The address interpretation sub task involves interpretation of the address components for the task at hand viz, dispatch sorting and forwarding of mail or distribution sorting for delivery of mail.

The paper also outlines the automated address interpretation scheme employed for dispatch sorting and forwarding of mail. A computational methodology for validation of destination postal addresses and automatic sorting of mail for dispatch at a mail sorting office especially in the Indian context is presented. The methodologies presented are modeled around the expert sorters, who automatically correct the addresses and sort the mail to a proper destination. It employs a frame representation for the knowledge base deployed for the purpose. The address validation technique also finds application for bulk mailers in ascertaining the veracity of the addresses obtained through different means such as feedback forms.

The paper is organized into 5 sections. The section 2 provides literature survey pertaining to postal automation with emphasis on postal address validation. Section 3 presents the knowledge based automated solution to mail sorting. It also describes the knowledge base and the rule base employed in the work. Section 4.0 presents the experimental results and their analysis. Section 5.0 gives the conclusion. Appendix-1 gives the fast search technique used in this system.

2.0 Literature Survey

There is a spurt of activity in postal automation area in recent times [1-4]. Nagabhushan (1998) identifies the different aspects of postal services that need to be automated. The literature survey reveals that

researchers around the world are addressing various issues related to building various subsystems required for postal automation. NEC in Japan is promoting postal automation research in a big way [1,7,8]. Specialized groups are working towards tackling various issues in postal automation at Centre for Excellence in Document Analysis and Recognition (CEDAR) at Buffalo University, USA and Stanford Research International (SRI). The Department of Studies in Computer Science, University of Mysore, India is also actively involved in tackling postal automation issues in the Indian Context

Premalatha and Nagabhushan, (2001), present an algorithmic prototype for automatic verification and validation of postal addresses. Nagamani and Nagabhushan, (2003) discuss a knowledge based approach to generation of destination postal codes from the postal addresses after validation. The different aspects of postal mail processing are discussed in [7,8]. Srirangaraj Setlur et. al (2001) propose a truthing, testing and evaluation mechanism for postal addresses, which validates the address and encodes it for speedier sorting using United States Postal Service (USPS) data. Address validation is an essential component for automatic sorting of postal mail, otherwise the mail sorting will be erroneous. The validation should be carried out by using the knowledge about destination places and their geography. The present work proposes a knowledge based solution to address validation and its use in automatic sorting of postal mail. Srihari et.al, (1999), present a formal method for information theoretic analysis of postal address components, which helps in their validation. The postal automation fraternity are working towards development of techniques for automation of various sub tasks of mail processing/ handling.

3.0 The Knowledge based Address Validation and Sorting System

Postal mail sorting involves reading/ deciphering of the destination address and classifying them into one of the postal bags, which are forwarded to other sorting offices/ post offices for further sorting or delivery. The proposed system involves development of techniques for address validation and address interpretation for mail sorting based on the local knowledge and the predefined procedure for mail forwarding in the dispatch sorting office. The address validation technique checks for consistency among address components and corrects them, if necessary. Address validation generally verifies the destination address written on the mail for its consistency. Every mail specifies the destination area and place by providing area and place names as well as PIN code. The validation process checks for the correctness of the place and area names and their mapping to PIN is verified. The system further corrects the address if there is inconsistency, particularly if there is inconsistency in the information conveyed by the area/place name and PIN code the former is assumed to be correct and proper PIN correction is carried out. If the mail does not contain PIN and the other address components are validated then the PIN is generated to the extent possible depending on the information available at the sorting office. The system works on the premise that the destination place and area are probably more correct than the PIN code. The validated/corrected address is then employed for mail dispatch sorting at the dispatch sorting office. The proposed mail sorting system is modeled on the current functionality of a sorting office.

The knowledge based address validation and sorting system processes the input address components and validates/ corrects them for further sorting of postal mail. Figure-3 gives the system block diagram of the proposed knowledge based fast validation and sorting system. The system comprises of two processes namely address validation and mail sorting. The address validation module checks for consistency among the various address components, whereas the mail sorting module sorts the mail using the rule base employed by the mail sorting office.
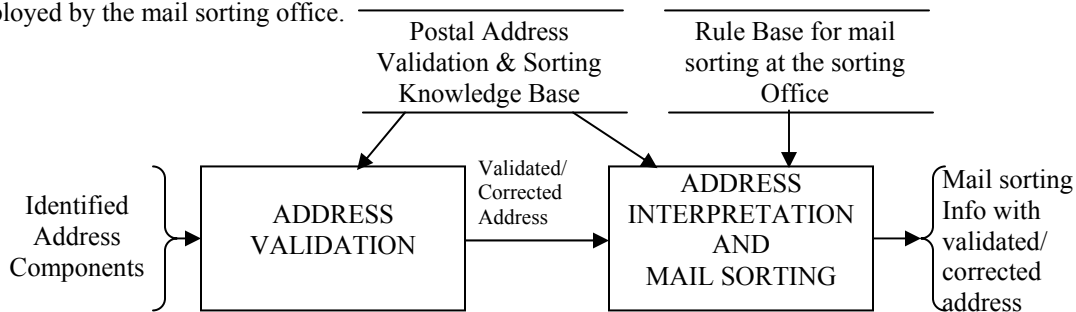
Figure-3: The Knowledge based Address Validation and Sorting System

The proposed knowledge based system makes a few assumptions as detailed below.

- The address validation and mail sorting modules make use of the following named address components for dispatch sorting of postal mail.
  - o Area Name
  - o Place Name
  - o PINCODE
  - o District Name
  - o State Name
  - o Country Name.
  - o Letter Type (For example Post card, inland letter etc)
- The place name/ area name/ district name can be correctly spelt or mis-spelt or abbreviated, also an alias can be used. Some of the examples are as follows,
  - o Bangalore: Correctly Spelt
  - o Bangaloor:Mis- Spelt
  - o B'lore: Abbreviated form
  - o Madras/Chennai: Alias
- The PIN if it exists may contain only 1 digit (the last digit of PIN), two digits (the last two digits) or 6 digits or it might be wrongly specified. For example, the PIN for Bangalore may be specified in the following ways.
  - o Bangalore-1 : One digit
  - o Bangalore-01: Two digits
  - o Bangalore-560 001: Six digits
  - o Bangalore-570 001: Wrong PIN
- In case of ambiguity between place name and PIN, place name is always assumed to be correct in this work.

It is required to store the geographic and other relevant knowledge about places at the mail sorting centre to accomplish automatic sorting of postal mail. The structure of the knowledge base employed is explained in section 3.1. The rule base used for sorting is described in section 3.2.

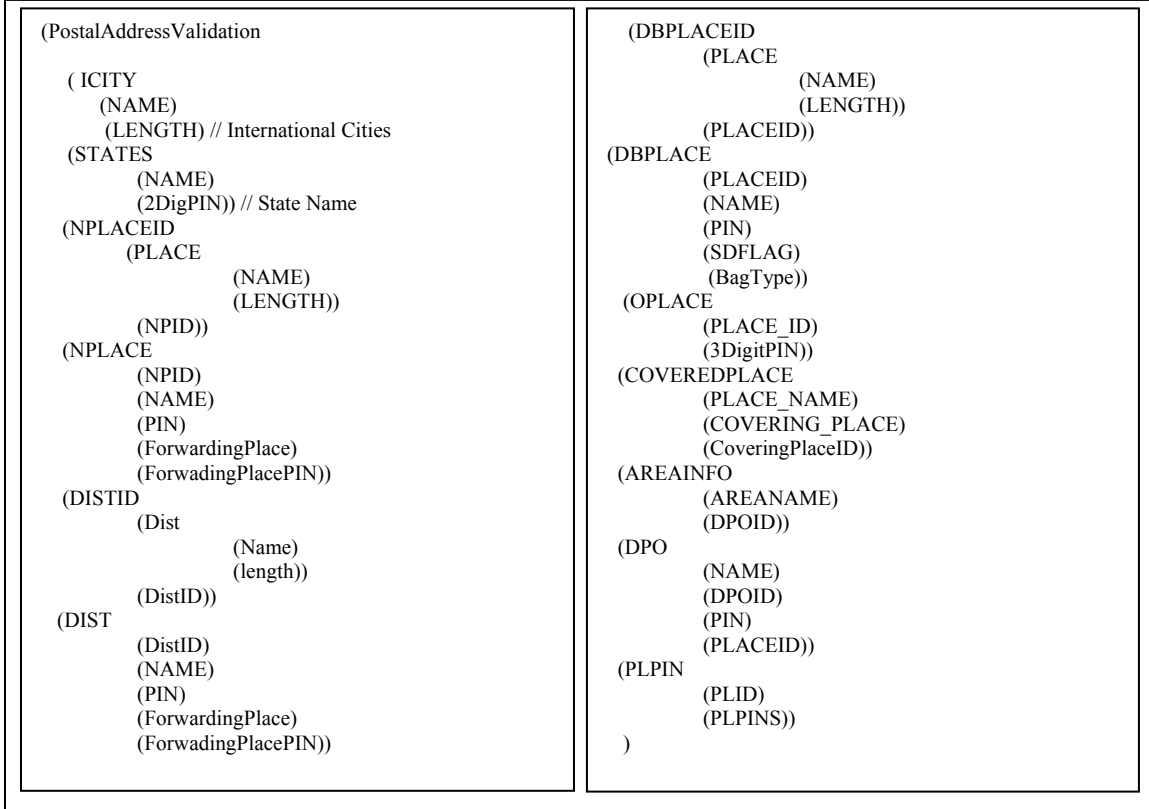## 3.1 The Postal Knowledge Base for dispatch sorting

The information required for address validation such as various forms of place names, area names, PIN codes etc, are to be stored in the knowledge base. A frame structured knowledge base has been devised for the purpose. Frames are used to represent knowledge and are general record like structures which consist of a collection of slots and slot values. The slots may be of any size and type and can take any number of values. Slots typically have names and values or sub fields called facets. Facets may also have any number of values [Dan W Patterson, 1997]. This property of the frames makes it suitable for representing the postal knowledge required for validation.

The knowledge base has been designed considering the destination addresses of about 1000 real postal addresses, which are taken as representative samples and also the mandated list of places and procedures in the sorting offices of India Posts. The developed knowledge base includes information about various forms of place names, area names, PIN codes, places to which direct bags are closed, known cities / places, the districts within the postal circles of the sorting district, the covered places (those places covered by a place to which direct bags are closed) and the like. The structure of the knowledge base is depicted in Figure-4. The content of the knowledge base is sorting office dependant, but the proposed structure can be employed by any sorting office. The present system uses the knowledge base developed for Bagalkot sorting office. Table 1.0 gives a list of four instances of every slot of the postal address validation frame.

3.2 The Rule Base for dispatch sorting
Every sorting office affiliated to India Posts is attached to a postal circle. There are 19 postal circles in India. (Refer www.indiapost.org for details). The postal circle mandates a list of places to which a direct bag is to be closed from a given sorting office based on study of mail pattern. Using this mandated list along with the type of the bag to be closed the sorting office sorts the mail to appropriate bags. The list contains every sorting district within the postal circle, every delivery post office within the sorting district, some delivery post offices outside the sorting district to which a lot of mail is destined, sorting offices of nearby places outside the circle to where a substantial quantity of mail is being sent etc.. The sorting is currently done manually by a set of sorting assistants using the guiding principle that the mail should not be held back and it should be on its way to the destination immediately. The India post has developed a set of rules for sorting, which are almost the same world over. The automatic sorting process is also to employ the

4

same set of rules. The rules employed by the developed system for sorting are listed in Figure -5. The mail sorting is a skilled job and the sorter needs to know the names of places and their status with respect to whether they are within circle, outside circle, if within circle in which sorting district etc. The automatic sorting process needs to imbibe this knowledge of the expert sorter. Hence the automatic process employs the knowledge base described in the previous section. The complete automatic sorting process is explained in section 3.3.

```
(PostalAddressValidation                          (DBPLACEID
                                                        (PLACE
   ( ICITY                                                    (NAME)
      (NAME)                                                  (LENGTH))
       (LENGTH) // International Cities                 (PLACEID))
    (STATES                                       (DBPLACE
          (NAME)                                        (PLACEID)
          (2DigPIN)) // State Name                      (NAME)
    (NPLACEID                                           (PIN)
        (PLACE                                          (SDFLAG)
                 (NAME)                                 (BagType))
                 (LENGTH))                       (OPLACE
          (NPID))                                       (PLACE_ID)
    (NPLACE                                             (3DigitPIN))
          (NPID)                                 (COVEREDPLACE
          (NAME)                                        (PLACE_NAME)
          (PIN)                                         (COVERING_PLACE)
          (ForwardingPlace)                             (CoveringPlaceID))
          (ForwadingPlacePIN))                   (AREAINFO
    (DISTID                                             (AREANAME)
        (Dist                                           (DPOID))
                 (Name)                          (DPO
                 (length))                              (NAME)
          (DistID))                                     (DPOID)
    (DIST                                               (PIN)
          (DistID)                                      (PLACEID))
          (NAME)                                 (PLPIN
          (PIN)                                         (PLID)
          (ForwardingPlace)                             (PLPINS))
          (ForwadingPlacePIN))                     )
```

<p align="center">Figure-4: The Postal Knowledge Base</p>

<p align="center">Table-1: Instances of the Knowledge Base</p>

| Sl N0	Frame Slot	Instance 1	Instance 2	Instance 3	Instance 4
1	ICITY	((BONN)(4))	((PERTH)(5))	((BAGHDAD)(7))	((PRETORIA)(8))
2	STATES	((ASSAM)(78))	((HARYANA) (12-13))	((KARNATAKA) (56-59))	((MANIPUR)(79))
3	NPLACEID	(((4)(KULU)) (001))	(((8)(NEWDELHI)) (005))	(((11) (PONDICHERRY)) (040))	(((12) (BHUVANESHWAR)) (042))
4	NPLACE	((001)(KULU) (175101) (BANGALORE) (56000))	((004)(COIMBATORE) (641001)(BANGALORE) (560001)	((010)(MUMBAI) (400001) (PUNE) (411001))	((005)(NEWDELHI) (111111) (BANGALORE) (560001))
5	DISTID	(((3)(BGK)) (018))	(((8)(BAGALKOT)) (018))	(((6)(MYSORE)) (006))	(((6)(HASSAN)) (004))
6	DIST	((001)(PUNE) (411001)(PUNE) (411001))	((002)(GADAG) (582101) (GADAGSTG) (582103))	((003)(KOCHI) (682001)(KOCHI) (682001))	((004)(HASSAN) (573201) (HASSANSTG) (573201))
7	DBPLACEID	(((3)(BJP)) (004))	(((5)(BEVUR)) (007))	((5)(KUMTA)) (016))	(((6)(BADAMI)) (001))

8	DBPLACE	((005)(PUNE) (L,M,D,R)(F))	((006)(BELUR) (L,M,D,R)(T))	((008)(BIRUR) (L,M,D,R)(F))	((009) (BANGALORE) (L,M,D,R)(F))
9	OPLACE	((BELGAUM) (003)(590))	((HUBLI) (012)(580))	((TELGI)(018) (586))	((PUNE)(005) (411))
10	COVEREDPLACE	((PATTADKAL) GUDUR)(011))	((NANJANGUD) (MYSORE)(028))	((MUCHKANDI) (BAGALKOT) (002))	((HALKURKI) (BADAMI)(001))
11	AREAINFO	((BAGALKOT UDYOG LTD)(004))	((BALULMATH BUILDING)(002))	((BANASHANKA RITEMPLE)(001))	((BAS ENGG COLLEGE)(003))
12	DPO	((010) (KERUR) (587206)(014))	((011)(BILGI (587116)(020))	((012)(HUNNUR) (587119)(023))	((013)(KONNUR) (587121)(024))
13	PLPIN	((004)(5861-01,02))	((005) (4110-01,02))	((006)(587114))	((007)(587115))

1. If mail is destined outside the country then sort and forward it to the air sorting office at the
    postal head quarters.
2. If the letter is destined outside the postal circle and is not to one of the nearby places outside
    postal circle to which direct bags are closed and the destination place is not covered by these
    places, then forward the letter to the circle head quarters for forwarding to the appropriate
    postal circle.
3. If the letter is destined within the postal circle and outside the sorting district then,
            i) if the letter is to a place, which has a direct bag to delivery post office, then sort and
            forward the letter to the corresponding delivery post office.
             ii) if the letter is to a place to which there is no direct bag but is within a sorting district,
            then forward the mail to the appropriate sorting district
 4. If the letter is destined within the sorting district then forward it the appropriate delivery post
    office.

Figure-5: The Rule base for Mail Sorting

3.3 Fast Address Validation and Sorting Technique

Automatic mail sorting requires to imbibe the features of manual sorting, especially the knowledge
employed by the expert sorters, which they have gained by experience. Basically the process of
deciphering the names of the places and their status with respect to the sorting rules is a skilled activity and
needs the use of knowledge. Also the addresses are written in many different forms using whatever
information is known about the destination/ addressee. Hence there is a need to ascertain what the
destination address really means and this is accomplished by address validation. The address validation
process checks for and removes inconsistency present in the destination address. For achieving address
validation and proper sorting the proposed system employs the knowledge base and the rule base described
in previous sections. The validation process begins by checking for country information, known cities
(international and national), then for state information etc. The outline of the validation and sorting process
is depicted in Figure-6 in the form of a flowchart
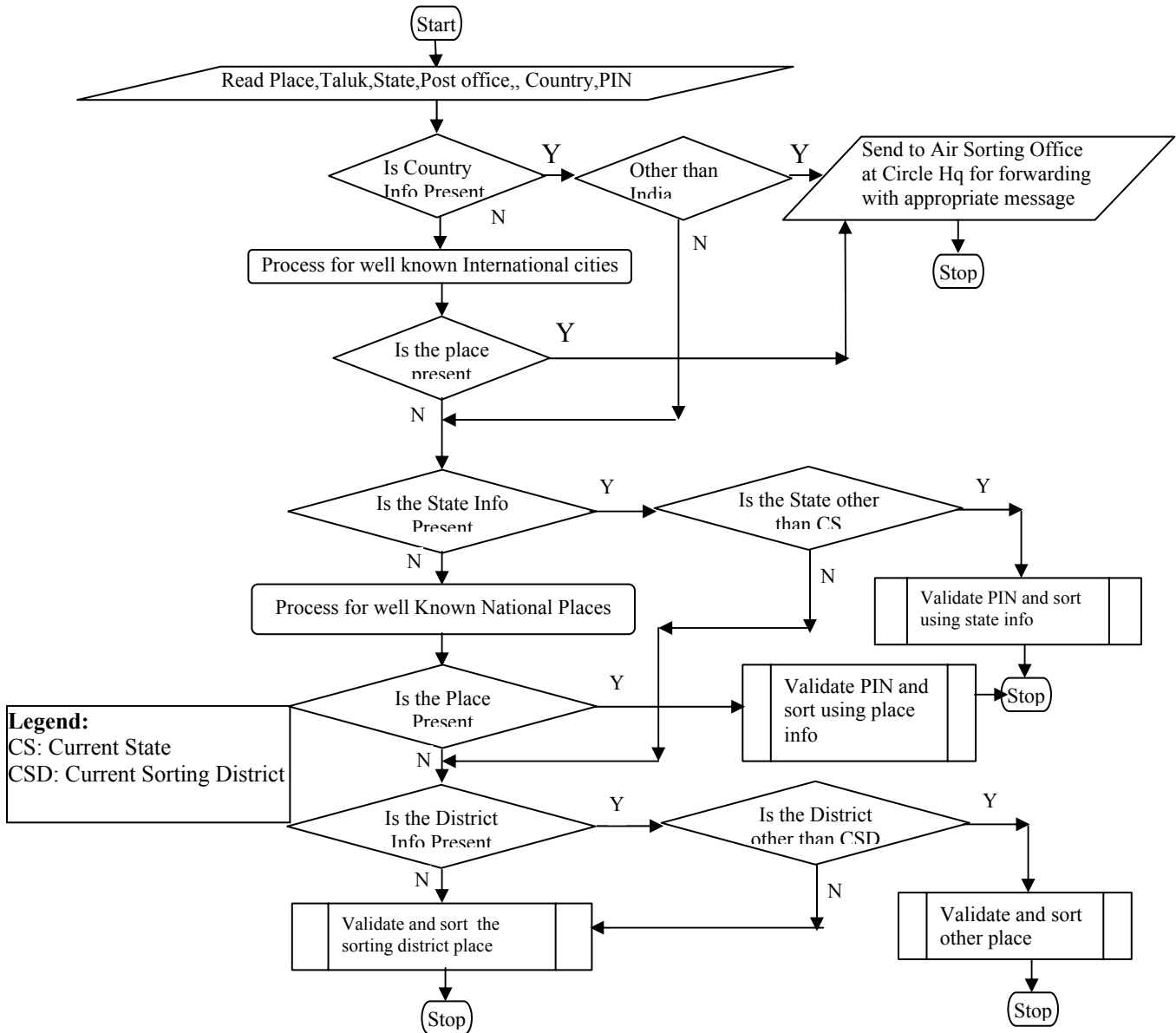
Figure 6: The overall Flow of Validation and Sorting at a Sorting Office

The proposed technique takes the computer recognizable form of destination address, validates the PIN and place to the extent possible, depending on the available information at the sorting office and specifies the place to which the mail is to be forwarded. The technique employs a fast search methodology for ascertaining the presence of place/ area / district in various forms in the knowledge base. The search technique is devised for searching in the frame structured knowledge base which is stored as a separate sorted flat file for information in each slot. In the new proposed search technique the length of the string is first compared and if it matches the first and last characters are compared for the match. If there is a match then a full comparison is made for ascertaining the match otherwise the next word is considered for the match. The search algorithm is detailed in Appendix-1.

4.0 Experimental Results and Analysis

The knowledge base for Bagalkot sorting office is built according to the representation technique discussed in section 3.1 and presented in Table-1. The validation and sorting system as presented in the flowchart of Figure-6 and incorporating the rule base of Figure-5, is implemented in C language and is tested on a P-III

machine with 128 MB RAM. The text strings forming the different address components are input to the system. The system outputs the validated addresses after correction, if necessary. Some of the mails may require manual intervention for validation and sorting. A sample set of test results are enlisted in Table-2. Most of the typical cases that are encountered in the sorting office are reported.

Table-2: The test results of validation for dispatch sorting

Sl. No.	Input Address	Output of the Algorithm	Remarks
1	Mr. S.G.Patil Near Sarnaiks House Extension Bagalkot-587102 Bagalkot Karnataka Type:Inland Letter	A:EXTENSION PL: BAGALKOT VPN: 587101 D: BAGALKOT S: KARNATAKA BT: L-BAG	The PIN code of the address is incorrect and is corrected by the algorithm. The mail goes to HPO Bagalkot. RT: FV
2	Dr. S.S.Gangoor Ayurvedic Medical College Badami-01 Bagalkot Karnataka Type: Envelope	A:AYURVEDIC MEDICAL COLLEGE PL: BADAMI VPN:587201 D:BAGALKOT S:KARNATAKA BT: L-BAG	The destination place is within sorting district and 6 digit PIN code is validated to the distribution area. The mail is forwarded to PO, Badami RT:FV
3	Mr. S.S.Angadi Vijay Medical Stores Extension Amingad-12 Bagalkot Karnataka, Type: Post card	A: EXTENSION PL:AMINGAD VPN:587112 D:BAGALKOT S: KARNATAKA BT: M-BAG	The area name is the same as in Bagalkot, still it is validated to distribution office. The mail goes to PO, Amingad. RT: FV
4	Shri G.G.Menisinakai Cloth Merchant Shastri Market Bjp-587102 Bijapur Karnataka, Type: Registered	PL: BIJAPUR VPN: 586 D: BIJAPUR S:KARNATAKA BT: R-BAG	The place is near but not in sorting district hence the address is validated upto forwarding sorting office, and sent to Bijapur sorting. The short form of place name is used here. RT: PV
5	Shri M.M.Patil Lakshmi Extension Gokak-591307 Belgaum Karnataka, Type: Letter	MFPL: LOKAPUR FPN: 591307 BT:TRANSIT BAG	The mails to Gokak are covered by Lokapur, hence the mail is sent to Lokapur. RT:PV
6	Mr. A.B.Waghmore Ghorpadi Poona-4 Maharashtra Type: Post card	PL:PUNE VPN:411001 D:PUNE S:MAHARASHTRA BT: M-BAG	The place is a nationally known place and hence the fully validated address is generated after correcting the pin code. RT:FV
7	Dr.S.S.Achar Dept of Mathematics UOM Manasagangothri Mysore 560 001 Karnataka Type: Cover	PL:MYSORE VPN:570 D:MYSORE S:KARNATAKA BT: D-Bag	The place is a far place to which a direct bag is closed but the given PIN is incorrect, the correct first 3 digits are generated and mail is forwarded to Mysore.
8	Mr Rajesh Fort area Belur-573115 Hassan Karnataka	Manual Intervention Required	As there is a place named Belur in Bagalkot district with code 587114, this mail is not validated manual intervention is required. RT: MI
9	Prof K Suresh Babu Reader National Institute of Technology Calicut Kerala Type: Registered	PL:KOZIKHODE VPN: 673001 S: KERALA FPL: KOCHI FPN: 682001 BT: R-BAG	As the place is a nationally known place, full validation is possible and the mail is forwarded to appropriate place RT: FV
10	Mr. Joseph A	MFPL: Air Sorting, Bangalore	As the destination place is a known

8

	Gilbey's Super Market Singapore. Type : Cover	BT: L-BAG	international place, the mail is forwarded to Air sorting post office at circle head quarters

Note: Table 2 uses some short forms which are described below.
- *'Type' in column 2 specifies the type of mail.*
- *Column 3 uses the following short forms*
  - *PL: Place*
  - *D: District*
  - *S: State*
  - *VPN: Validated PIN*
  - *MFPL: Forwarding Place*
  - *FPN : Forwarding PIN*
  - *BT: Bag Type in which the mail is forwarded*
- *Column 4 uses the following short forms*
  - *RT: Result Type*
  - *FV: Fully Validated*
  - *PV: Partially Validated*
  - *MI: Manual Intervention*

The system is exhaustively tested for sorting and validation for dispatch at Bagalkot sorting office and the results are summarized in Table-3. The results show that about 90.80% of the mail is either fully or partially validated and sorted by the proposed system, about 8.60% of the mail required manual intervention and 0.60% of the mail could not be resolved. Hence the methodology developed is sufficiently robust and can be used by any sorting office in India with appropriate knowledge base.

Table 3.0 : The results of implementation

Sl No.	Particulars	Observed Value	Percentage of total test mails
1	The number of mails tested	500	100
2	The number of mails completely validated	372	74.40
3	The number of mails partially validated	82	16.40
4	The number of mails that required human intervention	43	8.60
5	The number of mails that cold not be resolved	3	0.60

The search method employed by the algorithm is found to be an efficient technique for the proposed application as full comparison with all the information in the knowledge base is eliminated. The method assumes the use of established algorithms for address block location, character and numeral reading. The algorithm failed to properly validate the mail destined to a place outside the sorting district, but had a name similar to the name of a place within the sorting district and had incorrect pin code, such a mail required manual intervention. The places having similar names within a sorting district will generally be differentiated popularly by using some extensions, for example padasalgi and chikka padasalgi.

5.0 Conclusions

The knowledge based method for address validation and sorting of mail for dispatch sorting proposed in this paper is robust and is applicable to any sorting office. The methodology developed works in a manner similar to the current manual sorting operations, but does it more efficiently as the factors of fatigue and mistake are eliminated by the automatic process. The technique takes care of mis-spelt, correctly spelt, abbreviated names and also alias names of places and areas. The system validates or generates as many PIN digits as are possible using the information extracted from the mail destination address, with the help of the knowledge base and rule base.

When a network of sorting offices is established, the address validation and sorting methodology proposed here can be modified to completely validate the mail at the point of its origin and correctly sort it for speedy delivery. The success of the address validation and sorting system is largely dependant on the efficiency of the address reading system. The authors are working towards the development of various modules essential for postal automation. The system finds application at a sorting office which can be located at a place or attached to a train/ railway and it can also be used as an address validation system by bulk mailers, when the veracity of the address available is in question.

References

[1] Kanehiro Kubota and Kazunari Egami, (1999), " Technology trend of postal automation", NEC Research and Development, Vol 40, No. 2, Special Issue on Postal technology, pp 127-136

[2] Giovani Garibotto, 2002, "*Computer Vision in Postal Automation*" Elsag Bailey- TELEROBOT, 2002.

[3] P.Nagabhushan, 1998, " Towards Automation in Indian Postal Services : A Loud Thinking", Technovision , Special Volume, pp 128-139

[4] Rangachar Kasturi, Lawrence O'Gorman and Venu Govindaraju,2002, "Document Image Analysis: A Primer", Sadhana, Vol 27, Part 1, Febraury 2002, pp 3-22.

[5] M.R.Premalatha and P. Nagabhushan, 2001, "*An algorithmic prototype for automatic verification and validation of PIN code: A step towards Postal Automation*", Proceedings of AICTE-ISTE National Conference on Document Analysis & Recognition "NCDAR", 13th and 14th July 2001, pp 225-233

[6] M.R.Nagamani and P.Nagabhushan, 2003, "*Knowledge based approach to Determine the Destination Postal Code Through Address Block Extraction : A case study towards Postal Automation*", Proceedings of National Conference on Document Analysis and Recognition (NCDAR-2003) held at PESCE, Mandya, 11th and 12th July 2003, pp 152-163

[7] Masashige Iida, Hiroyuki Nagatadi and Tokuki Miura, 1999, "*Mail Automation System for the Overseas Market*", NEC Research & Development, Vol 40, No. 2 April 1999,Special Issue on Postal Technology, pp 192-197

[8] Yoshikazu Kobayashi, Nobuo Kuwagaki and Tokurou Hamada, 1999, " *Logistics Management System for the Mail Processing*", NEC Research & Development, Vol 40, No. 2 April 1999,Special Issue on Postal Technology, pp 187-191

[9] Srirangaraj Setlur, A Lawson, Venu Govindaraju and Sargur N Srihari,, 2001," Truthing, Testing and Evaluation Issues in Complex Systems", Sixth IAPR International Conference on Document Analysis and Recognition, Seattle, WA, pp 1205-1214

[10] Sargur N. Srihari, Wen-jann Yang and Venugopal Govindaraju, 1999, "Information Theortic Analysis of Postal Address Fields for Automatic Address Interpretation", Proceedings of Fifth International Conference on Document Analysis and Recognition (ICDAR-99), Bangalore India, pp 309-312

 [11] Universal Postal Union Address Standard, "FGDC Address Standard Version 2".

[12] Dan W Patterson,1997, "Principles of Artificial Intelligence", Prentice Hall International

Appendix-1

The Search Procedure

Input: L(length of WORD),WORD, FileName   **Output:** True/False

1.   Open the file with filename *"FileName"*
2.   Initialise FOUND to false.
3.   While not end of file do
    3.1.   Read the line of information(l,wd) // l is the length of the word and wd is the word
    3.2.   If (L=l) then
        3.2.1.   if (the first and last characters of the WORD and wd match) then
            3.2.1.1. if (theWORD and wd are same) then
                3.2.1.1.1.   Set FOUND to True and break
            3.2.1.2. Else
                3.2.1.2.1.   Continue
            3.2.1.3. Endif
        3.2.2.   Else
            3.2.2.1. Continue
        3.2.3.   Endif
    3.3. Else
        3.3.1.   Continue
    3.4. Endif
4.   End While
5.   Return Found
6.   Stop.

Note: All the files which are searched are organized with one entry on each line. Every line contains the length of the entry and the entry.