

Point Name Matcher



Submitted By: Refractions Research Inc.
Suite 400 - 1207 Douglas Street
Victoria, BC, V8W-2E7
Phone: (250) 383-3022
Fax: (250) 383-2140

Table of Contents

1	INTRODUCTION.....	3
2	POINT NAME MATCHER	3
2.1	INPUTS.....	4
2.2	OUTPUTS.....	5
3	MATCHING ALGORITHM.....	6
3.1	DEFINITIONS	6
3.2	POINT MATCHING.....	7
3.2.1	<i>Island Points</i>	<i>7</i>
3.2.2	<i>Waterbody Points</i>	<i>7</i>
3.2.3	<i>Matching Algorithm</i>	<i>8</i>
3.2.4	<i>Automated Matching Process.....</i>	<i>9</i>

Figures

Figure 1: Point Name Matcher Processes	3
Figure 2: Process Flow	11

Tables

Table 1: Islands Shapefile & BC Mainland Shapefile.....	4
Table 2: Toponymy Shapefile	4
Table 3: Named Points Shapefile	5
Table 4: List of Issues	5
Table 5: Definitions.....	6

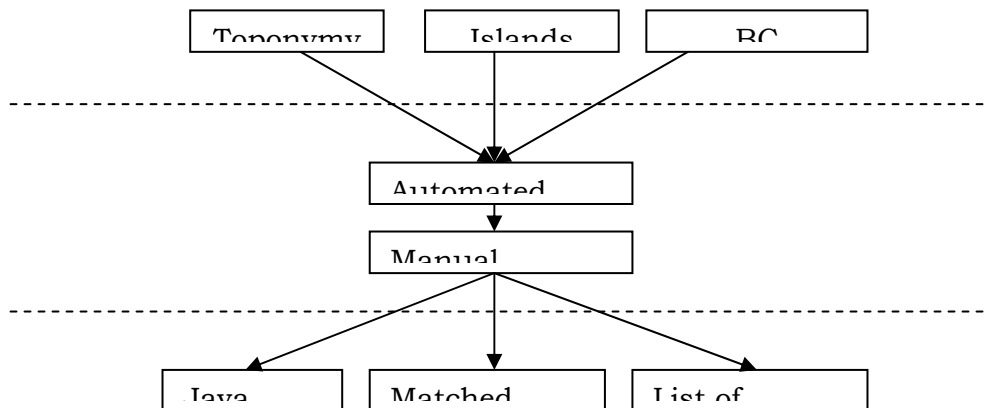
1 INTRODUCTION

The Point Name Matcher (PNM) tool allows a user to match a set of point toponyms to an island or coastline vertex (point). This tool includes an automated matching processes as well as a manual verification process where a user can verify and modify matches made during the automated matching process.

2 POINT NAME MATCHER

The following figure shows the stages for the Point Name Matcher process.

Figure 1: Point Name Matcher Processes



2.1 Inputs

The following tables describe the input files.

Table 1: Islands Shapefile & BC Mainland Shapefile

Field	Description
Shape	Polygon (in CW order)
Island Id	Unique id attached to each island polygon for identification purposes.
Island Type	0 – Saltwater Island 1 – Freshwater Island 2 – Saltwater & Freshwater Island (Example: Swishwashi Island)
Contained	ID of the island polygon that contains this polygon. Contained will be -1 if this island polygon is not contained within another island polygon. [This does not include the bc mainland polygon].

Table 2: Toponymy Shapefile

Field	Description
Point	Point (location of a letter)
TopoId	Toponym ID (all letters with the same toponym id form a toponym)
Elem_id	Position of letter in toponym name.
Letter	Letter associated with point.
Size	Size of the text.
Rotation	Rotation of the text.
GnisId	ID to link toponyms belonging to the same feature together.
Code	Code of the toponym: 1604 - Point
Code_Cat	Category classification of toponym.

2.2 Outputs

The following tables describe the output files.

Table 3: Named Points Shapefile

Field	Description
Shape	Point
Type	“Point”
Name	Name matched to point.
Gnisid	GnisId matched to point.
Island_ID	Island id associated with point. (For waterbody points this will be -1)
QAQC	QAQC flags associated with the match.

A comma-delimited file (.csv) will be produced of all the toponyms flagged with an issue by the user.

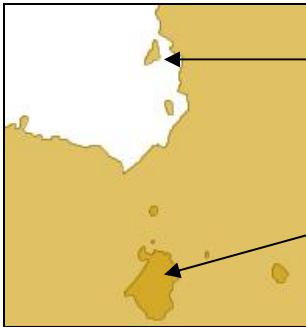
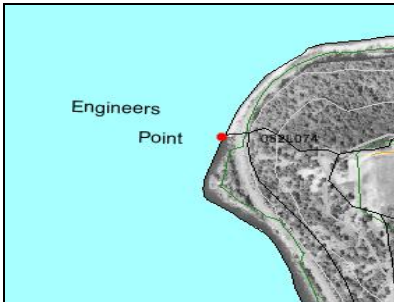
Table 4: List of Issues

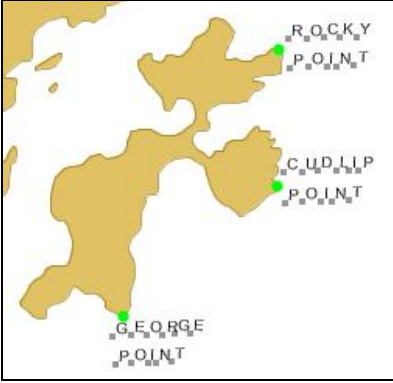
Field	Description
Toponym (Name)	Name of toponym to be split.
GNIS ID	Gnis ID for toponym to be split.
TOPO ID	Topo ID of toponym to be split.
COMMENTS	Comments about the issue entered by the user during the verification stage.

3 MATCHING ALGORITHM

3.1 Definitions

Table 5: Definitions

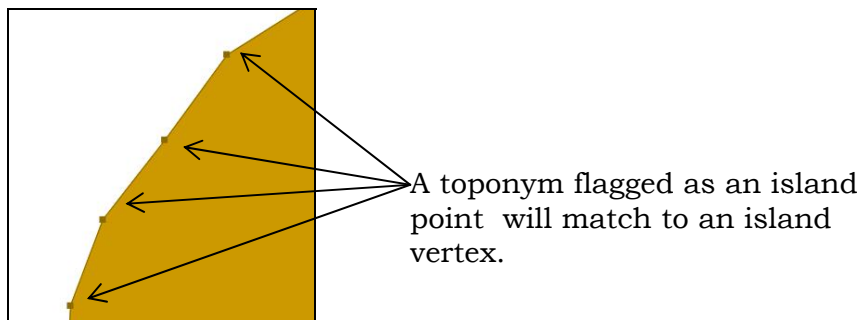
Toponym	A name on a map representing a feature. (<i>Example:</i> “Vancouver Island”)	
TopoID	An island feature can have more than one toponym. Each toponym is assigned a unique ID. (<i>Example:</i> “Vancouver Island” may appear on a map 10 times. Each of these toponyms would be assigned a separate, unique TopoID.)	
GnisID	All toponyms that represent the same feature are assigned the same GnisID. (<i>Example:</i> The 10 “Vancouver Islands” would all be assigned the same GnisID.)	
Contained Island	An island is contained if it exists inside another island (within a river, lake, or other water feature). <i>Example:</i> An island that is inside a lake on Vancouver Island would be a contained island. An island inside of a lake on the bc mainland would not be a contained island.  The map shows a large body of water (yellow) with a small island (yellow) inside it. An arrow points to the outer island with the label 'non-contained island'. Another arrow points to the inner island with the label 'contained island'.	
Waterbody Point	This is a point feature that is defined by a lake, river, or other waterbody.  The map shows a point feature labeled 'Engineers Point' on a shoreline. The point is marked with a red dot. The waterbody is labeled 'Shuswap Lake'.	‘Engineers Point’ is a Waterbody Point. It is defined by a waterbody (Shuswap Lake) not an island.

Island Point	This is a point feature defined by an island.  'Rocky Point', 'Cudip Point', and 'George Point' are all Island Points (associated with points on an island edge).
--------------	--

3.2 Point Matching

3.2.1 Island Points

Island points are matched to an island vertex near the toponym.



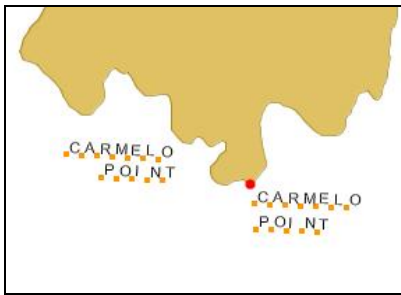
3.2.2 Waterbody Points

Waterbody points are not matched to island vertexes. Instead they are matched to a point in space that approximates the location of the point. This is done during the manual verification stage using WMS layers. Waterbody points are flagged and will be matched to a vertex on the waterbody polygon at a later time. This approach to waterbody points was taken due to the size of the waterbody (river and lake) datasets.

3.2.3 Matching Algorithm

The matching algorithm contains the following steps:

1. Load toponym, island, and bc mainland datasets from the corresponding shapefiles.
2. Create toponyms from raw dataset. The toponym shapefile contains one line per letter per topo id. At this stage the letters are combined (using the topo id) to generate a single toponym including all the letters associated with a topo id.
3. Perform the automated matching process.
4. Group toponyms based on GnisIDs. At this stage the toponyms and their matches are grouped together based on the toponym's GnisID.
 - In cases where two or more toponym are grouped together the grouped toponym is matched to the point with the highest rating from the automated matching step.



In this example there are two “Carmelo Point” toponyms (with the same GnisID). At this stage these two toponyms (and their matches) are joined to form a single grouped toponym. This is done so that toponyms with the same GnisID cannot be matched to different points.

Note: Toponyms may have the same name, but different GnisIDs. These toponyms are not joined as they reference different features.

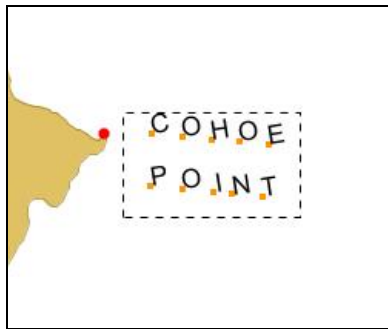
5. At this stage the automated matching is finished. The points, grouped toponyms, and matches are displayed to the user for verification and modification as necessary.

3.2.4 Automated Matching Process

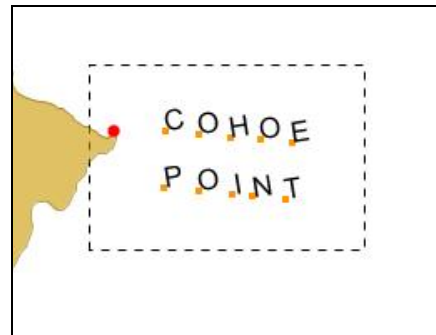
The automated matching process is described below.

For each toponym:

1. Find all edges that occur within a certain distance from the toponym.
2. If no edges are found in step (1) the distance is increased and the edges are searched again.



The initial search for edges results in no edges being found. Therefore the search area is expanded (right). This expanded search area matches multiple island edges.

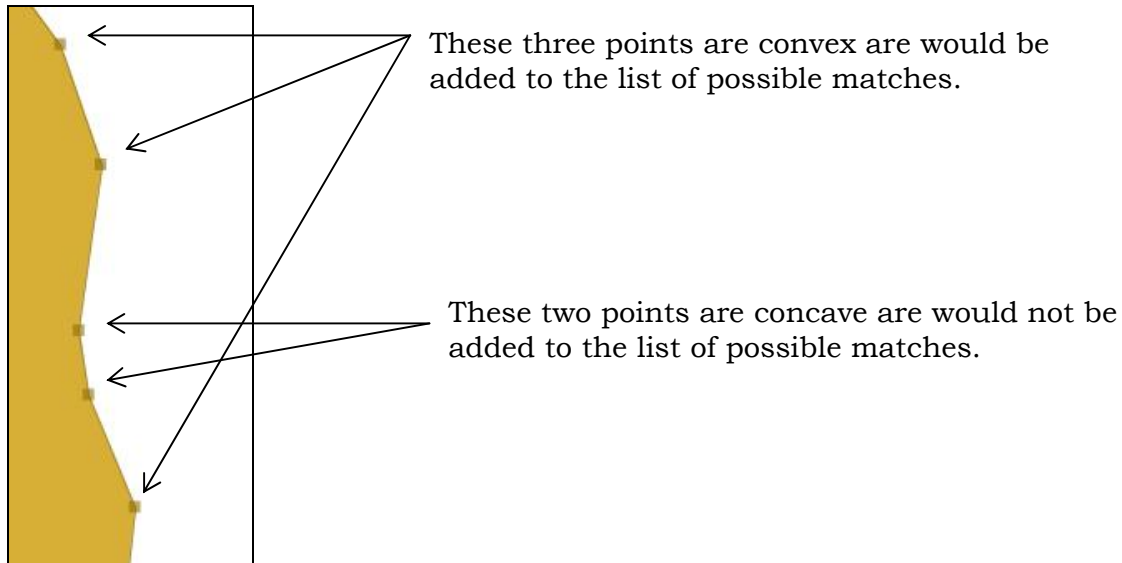


3. If no edges are found in step (2) then the point is flagged as a waterbody point and the matching process moves to the next toponym.



This particular toponym is inside the bc mainland polygon. As seen by the picture there are no island edges near the toponym; therefore the toponym is flagged as a waterbody polygon. During the manual verification process the user will be able to match the toponym to a point in space approximating the point location.

4. From all the vertices associated with the edges found from the first two steps a list of possible matches is created. A vertex is added to the list of possible matches if the angle at the vertex is convex. In this step we also find the largest island (by area) that is associated with one of the possible vertex matches.



5. If no possible matches are found then the point is flagged as a waterbody point and the matching process moves to the next toponym. *Otherwise*, each possible matches is rated and the best possible match is found. The rating factors include:
 - If the island associated with the vertex matches the largest island by area (found in previous step)
 - Total distance from each letter in the toponym to the vertex.
 - Angle of the vertex (more convex points will have a higher rating)
6. The match is set to the best point found, and the process moves to the next toponym.

Figure 2: Process Flow

